

STIC-Biotech/ChemLib

100781

Fr m: Baum, Stuart  
Sent: Friday, August 08, 2003 4:35 PM  
To: STIC-Biotech/ChemLib  
Subject: Sequence Search....thank you

Please do a standard and interference sequence search for:

- (1) SEQ ID NO:78
- (2) DNA's that encode SEQ ID NO:78

EXAMINER # 78896  
DATE:8/8/2003

SERIAL NUMBER:09/921992

MAIL BOX ROOM: CM1 9E12  
OFFICE ROOM # CM1 9D06

RESULTS FORMAT: PAPER

THANKS...Stuart

Stuart F. Baum Ph.D  
Art Unit 1638  
CM1 9D06  
(703) 305-6997  
stuart.baum@uspto.gov

RECEIVED  
AUG - 8 2003  
(STIC)

Edward Hart  
Technical Info. Specialist  
STIC/Biotech  
CMI 6B02 Tel: 305-9203

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 8/12/03  
Date Completed: 8/15/03  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: 141  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: 01/02  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_





# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 100781**

**TO: Stuart Baum**  
**Location: CM-1/9D06/9E12**  
**Art Unit: 1638**  
**Friday, August 15, 2003**

**Case Serial Number: 09921992**

**From: Edward Hart**  
**Location: Biotech-Chem Library**  
**CM1-6B02**  
**Phone: 305-9203**

**edward.hart@uspto.gov**

### **Search Notes**

Examiner Baum,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart







# STIC SEARCH RESULTS

## Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher* or contact:

Mary Hale, Information Branch Supervisor  
308-4258, CM1-1E01

## Voluntary Results Feedback Form

➤ I am an examiner in Workgroup:  Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature  
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library CM1-Circ-Desk



\_\_\_\_\_

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

Run on: August 12, 2003, 09:53:55 ; Search time 85 Seconds  
(without alignments)  
694.662 Million cell updates/sec

Title: US-09-921-992-78

Sequence: 1 MHNQAPIQRKSTRIVGNV.....RAKASQLDEARRIDVQVEK 372

Scoring table: BLOSUM62

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Database : A.Gensemseq\_19Jun03.\*

1:	/SIDS1/gcgdata/gensemseq/gensemseq-emb1/AA1980.DAT.*
2:	/SIDS1/gcgdata/gensemseq/gensemseq-emb1/AA1981.DAT.*
3:	/SIDS1/gcgdata/gensemseq/gensemseq-emb1/AA1982.DAT.*
4:	/SIDS1/gcgdata/gensemseq/gensemseq-emb1/AA1983.DAT.*
5:	/SIDS1/gcgdata/gensemseq/gensemseq-emb1/AA1984.DAT.*
6:	/SIDS1/gcgdata/gensemseq/gensemseq-emb1/AA1985.DAT.*
7:	/SIDS1/gcgdata/gensemseq/gensemseq-emb1/AA1986.DAT.*
8:	/SIDS1/gcgdata/gensemseq/gensemseq-emb1/AA1987.DAT.*
9:	/SIDS1/gcgdata/gensemseq/gensemseq-emb1/AA1988.DAT.*
10:	/SIDS1/gcgdata/gensemseq/gensemseq-emb1/AA1989.DAT.*
11:	/SIDS1/gcgdata/gensemseq/gensemseq-emb1/AA1990.DAT.*
12:	/SIDS1/gcgdata/gensemseq/gensemseq-emb1/AA1991.DAT.*
13:	/SIDS1/gcgdata/gensemseq/gensemseq-emb1/AA1992.DAT.*
14:	/SIDS1/gcgdata/gensemseq/gensemseq-emb1/AA1993.DAT.*
15:	/SIDS1/gcgdata/gensemseq/gensemseq-emb1/AA1994.DAT.*
16:	/SIDS1/gcgdata/gensemseq/gensemseq-emb1/AA1995.DAT.*
17:	/SIDS1/gcgdata/gensemseq/gensemseq-emb1/AA1996.DAT.*
18:	/SIDS1/gcgdata/gensemseq/gensemseq-emb1/AA1997.DAT.*
19:	/SIDS1/gcgdata/gensemseq/gensemseq-emb1/AA1998.DAT.*
20:	/SIDS1/gcgdata/gensemseq/gensemseq-emb1/AA1999.DAT.*
21:	/SIDS1/gcgdata/gensemseq/gensemseq-emb1/AA2000.DAT.*
22:	/SIDS1/gcgdata/gensemseq/gensemseq-emb1/AA2001.DAT.*
23:	/SIDS1/gcgdata/gensemseq/gensemseq-emb1/AA2002.DAT.*
24:	/SIDS1/gcgdata/gensemseq/gensemseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result	Score	Query Match	Length	DB	ID	Description
1	1866	100.0	372	22	AA045692	E. coli gcpE prote
2	1866	100.0	372	23	AAE19653	Escherichia coli G
3	868	46.5	368	23	AB047958	Listeria monocytog
4	814.5	43.6	403	23	ABP65322	Bifidobacterium id
5	794	42.6	359	19	AAU98328	H. pylori GHP0.76
6	785.5	42.1	415	22	AAU56212	Propionibacterium
7	761	40.6	378	22	AA091951	glutamicum protec
8	751.5	40.3	387	22	AA081205	Mycobacterium tube
9	751.5	40.3	387	24	ABP57500	Mycobacterium tube

10	534	28.6	421	24	ABP78766
11	490	26.3	611	20	AAI37716
12	470.5	25.2	621	20	AAI34971
13	462	24.8	740	23	AAE19650
14	460	24.7	679	22	AAE19653
15	446	23.9	752	21	AAE09883
16	445.5	23.9	603	23	AAE19652
17	445.5	23.9	686	23	AAE19651
18	428.5	23.0	165	21	AAE41371
19	428.5	23.0	169	23	ABP32470
20	398	21.3	144	23	ABU52039
21	242	13.0	776	22	ABG25879
22	119	6.4	113	21	AAE08755
23	119	6.4	113	21	AAE08752
24	113	6.1	489	22	AAE36357
25	111.5	6.0	578	20	AAE89273
26	110	5.9	108	21	AAE08756
27	110	5.9	108	21	AAE45733
28	109	5.8	256	22	AAU53170
29	106.5	5.7	578	19	AAE82392
30	105.5	5.7	578	19	AAE82355
31	105.5	5.7	578	21	AAI78566
32	105.5	5.7	578	23	AAE93408
33	105.5	5.7	578	23	AAU73205
34	105	5.6	545	23	AAU76187
35	104	5.6	605	23	ABE48093
36	102.5	5.5	542	23	ABE53701
37	102	5.5	546	23	ABP26321
38	102	5.5	546	23	ABP28434
39	101	5.4	338	23	ABE49213
40	101	5.4	523	22	AAE1619
41	100	5.4	343	19	AAE71467
42	100	5.4	343	19	AAE99880
43	100	5.4	502	22	AAE99993
44	99	5.3	541	21	AAI91970
45	99	5.3	541	22	AAU34139

## ALIGNMENTS

XX	RESULT 1
XX	ABAB45692
XX	ID ABAB45692 standard; Protein: 372 AA.
XX	AC
XX	ABAB45692;
XX	DT 15-MAR-2001 (first entry)
XX	DE E. coli gcpe protein.
XX	FW Isoprenoid; gcpe; yfjg; antimicrobial; transgenic plant; agriculture
XX	KW antimicrobial; antiparasitic; antiviral; fungicidal; herbicidal.
XX	OS Escherichia coli.
XX	PN W0200072022-A1.
XX	PD 30-NOV-2000.
XX	PF 20-MAY-2000; 2000MO-EP04592.
XX	PR 21-MAY-1999; 99DE-1023567.
XX	PR 21-MAY-1999; 99DE-1023568.
XX	PA (JOMA/) JOMAA H.
XX	PI Jomaa H;
XX	DR WPI: 2001-025196/03.
XX	DR N-PSDB: AAC82653.
XX	TT Incorporating gcpe and yfjg genes into viruses and cells, for

PT increasing isoprenoid content and identifying e.g. antimicrobial  
 agents, comprises using DNA sequences from bacteria or parasites  
 XX  
 XX  
 PS Disclosure; Page 15-17; 36pp; German.

CC This invention describes a novel method for incorporating GCPE and yfgB  
 CC genes into viruses and cells for increasing isoprenoid content and  
 CC identifying e.g. antimicrobial agents, comprises using DNA sequences (I)  
 CC from the gcpe or yfgB genes of bacteria or parasites or DNA sequences  
 CC (II) which hybridize to the specified genes or encode a plasmid protein  
 CC with the same biological activity as those encoded by the genes. The  
 CC invention also describes (2) plant cells containing (I) or (II); (2)  
 CC transformed plant cells, and transgenic plants regenerated from them,  
 CC that contain (I) or (II); (3) determining the enzymatic activity of a  
 CC GCPE protein; or (4) screening compounds (A) that have antimicrobial,  
 CC antiparasitic or antiviral activity in humans or animals or antiviral,  
 CC antifungal, fungicidal or herbicidal activity in plants. (I) and (II)  
 CC are used: (1) to increase the isoprenoid levels in viruses and cells;  
 CC (11) for determining the enzymatic activity of GCPE and yfgB proteins;  
 CC and (111) to identify compounds that inhibit activity of GCPE, i.e.  
 CC potential antibacterial, antimicrobial, antiparasitic or antiviral agents  
 CC for use in humans or animals, or antiviral, antiparasitic, fungicidal  
 CC or herbicidal agents for agriculture.

XX Sequence 372 AA;

Query Match 100.0%; Score 1866; DB 22; Length 372;

Best Local Similarity 100.0%; Pred. No. 7.6e-181; Mismatches 0; Indels 0; Gaps 0;

Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHNQAPIORRSTRIYGVNPIGDCAPLAVOSMTNRTTDEATVNOIKALERVGADIVR 60  
 DB 1 MHNQAPIORRSTRIYGVNPIGDCAPLAVOSMTNRTTDEATVNOIKALERVGADIVR 60  
 QY VSPPTMDAAEAFKLKQGVNPLVADHFDRIALKAFAEYDCLRINPNGINERIRM 120  
 DB VSPPTMDAAEAFKLKQGVNPLVADHFDRIALKAFAEYDCLRINPNGINERIRM 120  
 QY VSDCARDNIPRIGVNGSLSEKLDQERYGEPPTPALLSAMRHHVHDLRNLFDQFKSV 180  
 DB VSDCARDNIPRIGVNGSLSEKLDQERYGEPPTPALLSAMRHHVHDLRNLFDQFKSV 180  
 QY 121 VSDCARDNIPRIGVNGSLSEKLDQERYGEPPTPALLSAMRHHVHDLRNLFDQFKSV 180  
 DB 121 VSDCARDNIPRIGVNGSLSEKLDQERYGEPPTPALLSAMRHHVHDLRNLFDQFKSV 180  
 QY 181 KASDVFLLAVESYRLAKQIDOPHLGITEAGARGAVKSAIGLILSEGIGDRLRVS 240  
 DB 181 KASDVFLLAVESYRLAKQIDOPHLGITEAGARGAVKSAIGLILSEGIGDRLRVS 240  
 QY 241 AADPVEEIKVGFDIKSLRIRSGINFACPTCSQEPEDVIGTVNALBQRLEDDITPM 300  
 DB 241 AADPVEEIKVGFDIKSLRIRSGINFACPTCSQEPEDVIGTVNALBQRLEDDITPM 300  
 QY 301 STIGCVVNGPGEALVSEGVGTGNNKSGLYEDGVKRRDLNNDMDIDLEARIKASQ 360  
 DB 301 STIGCVVNGPGEALVSEGVGTGNNKSGLYEDGVKRRDLNNDMDIDLEARIKASQ 360  
 QY 361 EARRIDVOOVER 372  
 DB 361 EARRIDVOOVER 372

RESULT 2  
 AAEL19653  
 ID AAEL19653 standard; Protein; 372 AA.

XX AAEL19653;

XX 31-MAY-2002 (first entry)

DE Escherichia coli GCPE protein.

KM gcpe gene; methyl-D-erythritol phosphate; MEP; dimethylallyl diphosphate;  
 KM transgenic plant; isoprenoid compound; tocopherol; isopentenyl diphosphate;  
 KM food; feed source; transfection; single nucleotide polymorphism; SNP;  
 KM oxidative stress tolerance; UV tolerance; transformation; GCPE protein;

KM plant.

XX Escherichia coli.

XX WO200212478-A2.

XX 14-FEB-2002.

XX 06-AUG-2001; 2001WO-US24335.

XX 07-AUG-2000; 2000US-223483P.

XX (MONS ) MONSANTO TECHNOLOGY LLC.

XX Boronati A, Campos N, Rodriguez-concepcion M, Rohmer M, Seeman M;

XX Valentin HE, Venkatesh TV, Venkatramesh M;

XX WPI: 2002-227151/28.

XX N-PSDB: AAD31203.

PT gcpe nucleic acid which is an essential gene of the methyl-D-erythritol  
 PT phosphate pathway, encoding a fully defined GCPE protein which is  
 PT useful for increasing levels of tocopherol substrates in plants

XX Claim 1; Page 144-145; 155pp; English.

CC The invention relates to GCPE nucleic acid molecule, an essential gene  
 CC of methyl-D-erythritol phosphate (MEP) pathway that encodes rice,  
 CC Arabidopsis thaliana or Escherichia coli GCPE protein. GCPE is useful  
 CC for producing a transgenic plant such as Brassica campestris, B. napus,  
 CC canola, castor bean, coconut, cotton, crambe, linseed, maize, mustard,  
 CC oil palm, peanut, rapeseed, rice, safflower, sesame, soybean, sunflower,  
 CC or wheat with an increased isoprenoid (tocopherol) compound level. The  
 CC expression of GCPE protein in organisms increases the level of  
 CC tocopherol substrate such as isopentenyl diphosphate and dimethylallyl  
 CC diphosphate biosynthesis. Transgenic organisms overexpressing GCPE  
 CC protein can nutritionally enhance food and feed sources. Overexpression  
 CC of GCPE protein in transgenic plant may provide tolerance to stresses  
 CC e.g., oxidative stress tolerance such as to oxygen or ozone, UV  
 CC tolerance, etc. GCPE may be used to obtain nucleic acid molecules from  
 CC the same species, and to obtain nucleic acid homologues. GCPE is also  
 CC used as or primers. The recombinant vectors are used in plant  
 CC transformation or transfection. GCPE can also act as markers capable of  
 CC detecting polymorphisms such as single nucleotide polymorphisms (SNPs).  
 CC GCPE is also used to determine the level or pattern of expression of  
 CC the protein. The present sequence is Escherichia coli GCPE protein.

SQ Sequence 372 AA;

Query Match 100.0%; Score 1866; DB 23; Length 372;

Best Local Similarity 100.0%; Pred. No. 7.6e-181; Mismatches 0; Indels 0; Gaps 0;

Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHNQAPIORRSTRIYGVNPIGDCAPLAVOSMTNRTTDEATVNOIKALERVGADIVR 60  
 DB 1 MHNQAPIORRSTRIYGVNPIGDCAPLAVOSMTNRTTDEATVNOIKALERVGADIVR 60  
 QY VSPPTMDAAEAFKLKQGVNPLVADHFDRIALKAFAEYDCLRINPNGINERIRM 120  
 DB VSPPTMDAAEAFKLKQGVNPLVADHFDRIALKAFAEYDCLRINPNGINERIRM 120  
 QY 121 VSDCARDNIPRIGVNGSLSEKLDQERYGEPPTPALLSAMRHHVHDLRNLFDQFKSV 180  
 DB 121 VSDCARDNIPRIGVNGSLSEKLDQERYGEPPTPALLSAMRHHVHDLRNLFDQFKSV 180  
 QY 181 KASDVFLLAVESYRLAKQIDOPHLGITEAGARGAVKSAIGLILSEGIGDRLRVS 240  
 DB 181 KASDVFLLAVESYRLAKQIDOPHLGITEAGARGAVKSAIGLILSEGIGDRLRVS 240  
 QY 241 AADPVEEIKVGFDIKSLRIRSGINFACPTCSQEPEDVIGTVNALBQRLEDDITPM 300  
 DB 241 AADPVEEIKVGFDIKSLRIRSGINFACPTCSQEPEDVIGTVNALBQRLEDDITPM 300



CC given in ABP65258 to ABP66354 ligated in frame to a polynucleotide  
 CC encoding a heterologous polypeptide. (1) has antidiarrhetic and  
 CC antibacterial activities, and can be used as an inhibitor of *Salmonella*.  
 CC (1) (which is a probe) is useful for the detection and/or identification  
 CC of *Blifidobacterium longum* in a biological sample. A carrier containing  
 CC the lactic acid bacterium *Blifidobacterium longum* NCC2705 (NCIM 1-2618)  
 CC can be used for preventing and/or treating diarrhea brought about by  
 CC pathogenic bacteria and/or rotavirus. The carrier is a food composition  
 CC selected from milk, yogurt, curd, cheese, fermented milks, milk based  
 CC fermented products, ice-creams, fermented cereal based products, milk  
 CC based powders, infant formula, pet food or a pharmaceutical composition  
 CC selected from tablets, liquid bacterial suspensions, dried oral  
 CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.  
 CC (1) is useful in DNA arrays or chips to carry out analysis of the  
 CC expression of the *Blifidobacterium* gene. AB081844 to AB081850 represent  
 CC *Blifidobacterium* related nucleotide sequences given in the Sequence  
 CC Listing from the present invention but not mentioned further within the  
 CC specification.  
 CC N.B. The sequence data for this patent is not represented in the printed  
 CC specification but is based on sequence information supplied by the  
 CC European Patent Office.

XX Sequence 403 AA;

Query Match 43.63% Score 814.5; DB 23; Length 403;  
 Best Local Similarity 46.33% Pred. No. 8.4e-74;  
 Matches 167; Conservative 70; Mismatches 111; Indels 13; Gaps 4;

QY 3 NQAPIQ-RRKSTRIVGNVPIGDCAPITAVOSMTTPTDVAATVNOIKALERVADIVRV 61  
 DB 22 SESPLHRRKRSRRIMVGPVPGAPISVOSMTTTLTANPATLQQLAETLAACCDIVRV 81  
 QY 62 SVPTMDAAEAFKLIKQOVNPLVADIHFDRIALKAVEYGVDCILRPNIGNEERIRM-120  
 DB 82 AVPEQDDADALPEICRSPPIVADIHFOKTYVQALDAGAAVRVNGNIRKFDVGP 141  
 QY 121 VVDCARKNPIRIGVNSLEKDLQKYEPTPOALLSAMRVHDLRLNFOQFVSV 180  
 DB 142 ICKAATAGISLRIGVNSGLDKELVAKYGGPPEALVSAALKEAHMFEDVGFDFKISV 201  
 QY 181 KASVFLAVESYRLAIGDIOPLHLGITEAGGARSVAISGLISLSEGISGLTVSL 240  
 DB 202 KHHVIVTWEYRLLASRGWPLHLGVTEGAPWOGTISKSLAFGALLASISGITIRSL 261  
 QY 241 AADVEEIKYGFDLKSLRIRSGINFLACPTCSROEFDVIGTVALBORELDITPMV 300  
 DB 262 SAPAEVYKVCCKLEKGLRKRKFDLISPCSGRAGVDVQLASVTEGLKDVTAIRV 321  
 QY 301 SIICGVNNGPEALVSTEGVTGKNGKSGLYEDG-----VKRDLN-----NDMIDOLE 349  
 DB 322 AVMCICVNGPEAREADIGVASGNGKQIFIKGVIKTVPEDQIVDELITANDIAOME 381  
 QY 350 A 350  
 DB 382 A 382

RESULT 5  
 AAM98298  
 ID AAM98298 standard; Protein: 359 AA.

XX AAM98298;

XX 31-MAR-1999 (first entry)

XX H. pylori GHPD 76 protein.

XX GHPD protein; *Helicobacter* infection; gastroduodenal disease; gastritis;  
 XX peptic ulcer disease.

XX *Helicobacter pylori*.

XX W09843478-A1.

XX 08-OCT-1998.  
 PD 98WO-US06371.  
 XX 01-APR-1998;  
 PF 29-JUL-1997; 9705-0902615.  
 PR 01-APR-1997; 9705-083457.  
 PR 24-JUN-1997; 9705-0881227.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA (IMMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.  
 XX Al-Garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;  
 PI WPI: 1998-542293/46.  
 DR N-PSDB; AAX14017.  
 XX New isolated *Helicobacter* polynucleotides - used to develop products  
 PT for the diagnosis, prevention and treatment of *Helicobacter*  
 PT infections and gastrointestinal diseases  
 PS Claim 8; Page 372-374; 2054pp; English.

CC This sequence represents a *Helicobacter pylori* GHPD protein of the  
 CC invention. The polypeptides can be used for preventing or treating  
 CC *Helicobacter* infections, and gastroduodenal diseases associated with  
 CC these infections, including acute, chronic, and atrophic gastritis, and  
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be  
 CC used for the production of antibodies. The products can also be used for  
 CC detection and diagnosis.

XX Sequence 359 AA;

Query Match 42.6%; Score 794; DB 19; Length 359;  
 Best Local Similarity 45.8%; Pred. No. 8.5e-72;  
 Matches 162; Conservative 73; Mismatches 117; Indels 2; Gaps 2;

QY 9 RRKSTRIVGNVPIGDCAPITAVOSMTTPTDVAATVNOIKALERVADIVRVVPTMDA 68  
 DB 5 RVKTKQIFIGVAGDAPISVOSMTTPTDVAATVNOIKALERVADIVRVVAVSNEKD 64  
 QY 69 AEAFLIKQOVNPLVADIHFDRIALKAVEYGVDCILRPNIGNEERIRMVDCARDK 128  
 DB 65 ALALKEKLVSPPLIDHIFHFKFAL-IAQSVDAIRPNIGNEERIKAVVADACKK 123  
 QY 129 NIPRIGVNSGLEKDLQKYEPTPOALLSAMRVHDLRLNFOQFVSVKASVDELA 188  
 DB 124 NIPRIGVNSGLEKDLQKYEPTPOALLSAMRVHDLRLNFOQFVSVKASVDELA 182  
 QY 189 VESTYRLAKOTDOPHLGITEBAGGARSVAISGLISLSEGISGLTVSLAADPVEI 248  
 DB 183 IEAYRMLRPLVITYPFHGVTEAGNLFSSISKSAALAGLMEGIGDMRVSITGELENEI 242  
 QY 249 KYGFDLKSLRIRSGINFLACPTCSROEFDVIGTVALBORELDITPMVDSIIGCVN 308  
 DB 243 KVARAILRHSGRLKEGINWISCPGRIEAVLVMAIKVERKLSHITPTDISMGCVN 302  
 QY 309 GPGELAVSTLGVNGKNGKSGLYEDGVKRDLDNDMDIDOLEARIRASOLDEA 362  
 DB 303 ALGEAKHADMAIAGNRSGLITKEGKVIHKLAEKDLRETFVIEVENLAKEREKS 356

RESULT 6

AAU56212  
 ID AAU56212 standard; Protein: 415 AA.

XX AAU56212;

XX 27-FEB-2002 (first entry)

XX *Propionibacterium* acnes immunogenic protein #17108.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;

KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

XX W0200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US12865.

XX 21-APR-2000; 2000US-199047P.

XX 02-JUN-2000; 2000US-208841P.

XX 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORLXA CORP.

XX Skelly YW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L' Malsomeuve J, Zhang Y, Jen S, Carter D;  
 DR NPI; 2001-616774/71.

XX N-PSDB; AAS59573.

PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris.

XX Example 1; SEQ ID NO 17407; 1069pp; English.

CC Sequences AU93105-AU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 415 AA;

XX Query Match 42.1%; Score 785.5; DB 22; Length 415;

XX Best Local Similarity 44.2%; Pred. No. 7.9e-71;

XX Matches 159; Conservative 68; Mismatches 132; Indels 1; Gaps 1;

XX 9 RKRSTRIVGVNVPIDGAPIAVOSMTTTRTDVATNOIKALERVGADIVRVSPTMDA 68

XX 53 RKRTRQIVGVNVDLVGDAPIVSQMTTKTHDVGATLQIALATAGGCDIVRVAAPTMDK 112

XX 69 AEAFLIKQOVNVPVADIDHEDYRALKVAEYGVCLAINNGNIGN -EERIRAVVDCARD 127

XX 113 AEVLPITAKRSQIPYADIDHEDYRALKVAEYGVCLAINNGNIGN -EERIRAVVDCARD 172

XX 128 KNIPRIGVNVAGSLKEDLOEKYGEPTPOALLESMARHVDHRLNFDQFVSVASDVEL 187

XX 173 HGTSTRIGVNVAGSLKEDLOEKYGEPTPOALLESMARHVDHRLNFDQFVSVASDVEL 232

XX 188 AVESRRLAKQIDPLHIGTEAGAGSAGVSAIGLGLISEGIGDITRLVSLADPVEE 247

XX 233 MIRAEQIAKACDYPHLIGTEAGAGSAGVSAIGLGLISEGIGDITRLVSLADPVEE 292

QY 248 IKVFPILKSLRIRSGINFIACPTCSKQEPFVIGTVNALEORLEDTITPMDVSIIGCV 307

DB 293 VKVGIITLISLNRPRGLEIVSCPCRCQVDFLTIANVFTALLEGIDAPLRVAVGCV 352

QY 308 NGPGEALVSTLGVGNGKSGGLEYEDVRDRDNDNMIDQLERIKARASQDEARRIV 367

DB 353 NGLGEGREADLVGAAGNGKIFKKGIVRTVPEGEIVQFLVOEANRMADEMDTTCGAVEY 412

RESULT 7

XX AAG91951 standard; Protein; 378 AA.

XX AAG91951;

XX 26-SEP-2001 (first entry)

XX C glutamicum protein fragment SEQ ID NO: 5705.

XX Corynebacterium; amino acid synthesis; vitamin; saccharide;

XX organic acid synthesis.

XX Corynebacterium glutamicum.

XX EPI108790-A2.

XX 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 99JP-0377484.

XX 07-APR-2000; 2000JP-0159162.

XX 03-AUG-2000; 2000JP-0280988.

XX (KIOWA) KIOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 DR N-PSDB; AAH67170.

PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
 PT expression point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene

XX Claim 17; SEQ ID NO: 5705; 246bp + Sequence Listing; English.

CC The present invention provides a number of nucleotide and protein  
 CC sequences from the Corynebacterium bacteria Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of Corynebacterium glutamicum, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Corynebacterium glutamicum, and identifying a homologue of a gene derived  
 CC from Corynebacterium glutamicum. Corynebacterium bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.

XX Sequence 378 AA;

XX Query Match 40.8%; Score 761; DB 22; Length 378;

XX Best Local Similarity 44.3%; Pred. NO. 2.1e-68;

XX Matches 159; Conservative 64; Mismatches 124; Indels 12; Gaps 3;

XX 9 RKRSTRIVGVNVPIDGAPIAVOSMTTTRTDVATNOIKALERVGADIVRVSPTMDA 68

XX 5 RKRTRQIVGVNVDLVGDAPIVSQMTTKTHDVGATLQIALATAGGCDIVRVAAPTMDK 64

QY 69 AEAFLIKQOVNPLVADHFDYRIALKVAEYVDCLRNPNIGN-EERIMVVDCAARD 127  
 65 AEAFLIAKSPPIVADHFDYRIALDAGCAAVRPNINIKEDGKVKAKAAGD 124  
 QY 128 KNPIRIRGVNAGSLENDDEKYE-GEPTPOLLESAMRHVDHRLNFDQFVSKASDVE 186  
 Db 125 AGPIRIRGVNAGSLENDDEKYE-GEPTPOLLESAMRHVDHRLNFDQFVSKASDVE 186  
 QY 187 LAVESEYRLAKQIDOPHLITTEAGARSGAVSAIGLLSEBGIDTLRVSLADPVE 246  
 Db 185 LMWEATYRQLAEQSDYPIHLGVTEAGPRKFMGKIKSSVAFGALLSGIDITRVSLADPVE 244  
 QY 247 EIKVGPDIILKSLRIRSGINFIACPTCSROEPDYIGVNLAEQLEDTIIPMVSIIICV 306  
 Db 245 EIKVGPDIILKSLRIRSGINFIACPTCSROEPDYIGVNLAEQLEDTIIPMVSIIICV 304  
 QY 307 VNGGEALVSTLGTGNGKSGLYEDGVRKRDLDNNMIDOLEARIRAKASQIDEARRI 365  
 Db 305 VNGGEALVSTLGTGNGKSGLYEDGVRKRDLDNNMIDOLEARIRAKASQIDEARRI 365

RESULT 8  
 AAG81205  
 ID AAG81205 standard; Protein: 387 AA.

AC AAG81205;  
 DT 04-SEP-2001 (first entry)  
 DE Mycobacterium tuberculosis potential drug target protein SEQ ID 256.  
 KW Drug target; growth; organism viability; characterisation.

OS Mycobacterium tuberculosis  
 PN WO200135317-A1.  
 PD 17-MAY-2001.

PF 13-NOV-2000; 2000WO-US31152.  
 PR 12-NOV-1999; 99US-0165086.  
 PR 12-NOV-1999; 99US-0165124.  
 PR 01-FEB-2000; 2000US-0179534.

XX (REGC) UNIV CALIFORNIA.  
 PA  
 PI Eisenberg D, Rotstein SH, Marcotte EM;  
 DR WPI; 2001-329193/34.  
 DR N-PSDB; AAH52056.

PT Identifying nucleotide or polypeptide sequence for use as drug target,  
 PT involves providing algorithm that analyzes a functional relationship  
 PT between nucleotide or polypeptide sequences, and comparing the  
 PT sequences

PS Disclosure: Page 182; 207pp; English.

XX This invention relates to a method for identifying a nucleotide or  
 CC polypeptide sequence that may be a drug target, or essential for growth  
 CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092  
 CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium  
 CC tuberculosis proteins which are potential drug targets. The DNA and  
 CC protein sequences are used to illustrate the method of the invention. The  
 CC method involves providing an unknown nucleotide or polypeptide sequences,  
 CC and comparing it to a number of sequences along with at least one  
 CC algorithm capable of analysing a functional relationship between  
 CC nucleotide and polypeptide sequences. The method is useful for  
 CC characterising the function of nucleic acids and polypeptides that may be  
 CC useful as a target for a drug or essential for the growth or viability of  
 CC an organism.

SQ Sequence 387 AA:

Query Match 40.3%; Score 751.5; DB 22; Length 387;  
 Best Local Similarity 42.7%; Pred. No. 2e-67; Indels 11; Gaps 2;  
 Matches 156; Conservative 64; Mismatches 134;

QY 9 RRKSTRIVYGNVPIGDPAPIVAOVMTRTVEATVNOIKALERVGADIVRSVPTDA 68  
 Db 18 RRATROLAMGNVGVSDHVSVOVMCTKTHVYNSLQIALTLAAGCDIVRACPRQED 77  
 QY 69 AEAFLIKQOVNPLVADHFDYRIALKVAEYVDCLRNPNIGN-EERIMVVDCAARD 127  
 Db 78 ADALAEIAHSHSQIPVADHFDYRIALDAGCAAVRPNINIKEDGKVKAKAAGD 137  
 QY 128 KNPIRIRGVNAGSLENDDEKYE-GEPTPOLLESAMRHVDHRLNFDQFVSKASDVE 187  
 Db 138 AGPIRIRGVNAGSLENDDEKYE-GEPTPOLLESAMRHVDHRLNFDQFVSKASDVE 197  
 QY 188 AVESEYRLAKQIDOPHLITTEAGARSGAVSAIGLLSEBGIDTLRVSLADPVE 247  
 Db 198 MVAAYELIARCDYPIHLGVTEAGPRKFMGKIKSSVAFGALLSGIDITRVSLADPVE 257  
 QY 248 EIKVGPDIILKSLRIRSGINFIACPTCSROEPDYIGVNLAEQLEDTIIPMVSIIICV 307  
 Db 258 VKVGNVYLESILNRPRLSEIVSCPSGCRQAVDYTLANETAGLDGDLPLRVAVMGCV 317  
 QY 308 NGPGEALVSTLGTGNGKSGLYEDGVRKRDLDNNMIDOLEARIRAKASQIDEARRI 367  
 Db 318 NGPGEALVSTLGTGNGKSGLYEDGVRKRDLDNNMIDOLEARIRAKASQIDEARRI 367  
 QY 368 QOVER 372  
 Db 368 EMGEQ 372

RESULT 9  
 ABP57500  
 ID ABP57500 standard; Protein: 387 AA.

AC ABP57500;  
 DT 28-APR-2003 (first entry)  
 DE Mycobacterium tuberculosis protein SEQ ID NO:129.

DE Mycobacterium tuberculosis; infection; antibacterial; tuberculostatic;  
 KW immunostimulant; vaccine; gene therapy; mycobacterial infection.

OS Mycobacterium tuberculosis.

PN WO2003000721-A2.

PD 03-JAN-2003.

PF 21-JUN-2002; 2002WO-GB02845.

PR 22-JUN-2001; 2001GB-0015365.

PR 07-SEP-2001; 2001GB-0021780.

PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.

PI James BW, Bacon J, Marsh P;

DR WPI; 2003-201403/19.

DR N-PSDB; AB271126.

PT New mycobacterial peptide, its fragment, variant or derivative, useful  
 PT as vaccine for treating or preventing mycobacterial infections, and as  
 PT diagnostic reagents for identifying such infections

PS Claim 2; Page 235-236; 246pp; English.

XX ABP57436 to ABP57504 represent mycobacterial amino acid sequences (1)



RESULT 10
ABP78786
ID ABP78786 standard; Protein; 421 AA.
XX
XX AC
XX
XX
DT 07-MAR-2003 (first entry)
XX
DE N. gonorrhoeae amino acid sequence SEQ ID 4102.
XX
XX
KM Antibacterial; Infection; vaccine; gene therapy
XX
OS Neisseria gonorrhoeae.
XX
PN WO200279243-A2.
XX
XX
PD 10-OCT-2002.
XX
PF 12-FEB-2002; 2002WO-IB02069.
XX
XX
PR 12-FEB-2001; 2001GB-0003424.
XX
XX
PA (CHTR-) CHIRON SPA.
XX
XX
TI Fontana MR, Piazza M, Maignani V, Monaci E;

PD 10-JUN-1999.



DB 258 GIIRSAVIGITLLAEGIDTRISLNGCPTTEIPVCDLRRHKRI 302

## RESULT 13

AAE19650 standard; Protein; 740 AA.

AAE19650;

31-MAY-2002 (first entry)

Arabidopsis thaliana GCPE protein.

GCPE gene; methyl-D-erythritol phosphate; MEP; dimethylallyl diphosphate; transgenic plant; Isoprenoid compound; tocopherol; isopentyl diphosphate; food; feed source; transfection; single nucleotide polymorphism; SNP; oxidative stress tolerance; UV tolerance; transformation; chromosome V; plant; GCPE protein.

Arabidopsis thaliana.

Key Location/Qualifiers

Cleavage-site 32.33 /note="Plastidial targeting peptide cleavage site"

MO200212478-A2.

14-FEB-2002.

06-AUG-2001; 2001MO-US24335.

07-AUG-2000; 2000US-223483P.

(MONS) MONSANTO TECHNOLOGY LLC.

Boronat A, Campos N, Rodriguez-concepcion M, Rohmer M, Seeman M;

Valentin HE, Venkatesh TV, Venkatesh M;

WPI; 2002-227151/28.

GCPE nucleic acid which is an essential gene of the methyl-D-erythritol phosphate pathway encoding a fully defined GCPE protein which is useful for increasing levels of tocopherol substrates in plants

Claim 1; Page 140-142; 155pp; English.

The invention relates to GCPE nucleic acid molecule, an essential gene of methyl-D-erythritol phosphate (MEP) pathway that encodes rice, Arabidopsis thaliana or Escherichia coli GCPE protein. GCPE is useful for producing a transgenic plant such as Brassica campestris, B. napus, canola, castor bean, coconut, cotton, crambe, linseed, maize, mustard, oil palm, peanut, rapeseed, rice, safflower, sesame, soybean, sunflower, or wheat with an increased isoprenoid (tocopherol) compound level. The expression of GCPE protein in organisms increases the level of tocopherol substrate such as isopentyl diphosphate and dimethylallyl diphosphate biosynthesis. Transgenic organisms overexpressing GCPE protein can nutritionally enhance food and feed sources. Overexpression of GCPE protein in transgenic plant may provide tolerance to stresses e.g., oxidative stress tolerance such as to oxygen or ozone, UV tolerance, etc. GCPE may be used to obtain nucleic acid molecules from the same species, and to obtain nucleic acid homologues. GCPE is also used as or primers. The recombinant vectors are used in plant transformation or transfection. GCPE can also act as markers capable of detecting polymorphisms such as single nucleotide polymorphisms (SNPs). GCPE is also used to determine the level or pattern of expression of the protein. The present sequence is Arabidopsis thaliana GCPE protein. GCPE gene is located on chromosome V.

Sequence 740 AA;

Query Match 24.8%; Score 462; DB 23; Length 740;

Best Local Similarity 38.6%; Pred. No. 1 6e-37; Matches 105; Conservative 47; Mismatches 86; Indels 34; Gaps 5;

QY 9 RRRSTRIVYVGNPDIGGAPLAVQSMNTRTDTVEATVNOIKALERYGADIVRVSEPTMA 68  
 DB 84 RRRSTRIVYVGNVALGSEHPIRIQTMTSDPKITGVDEVMRIADGADIVRTVYQKKE 143  
 QY 69 AEAFA-----KLIRQVNVPLVADIHEDYRALKVAEYGDCLIRNGINERIRIMVYD 123  
 DB 144 ADACFEIKDLVOLNMYNPLVADIHFAFVLAFAEC-FDKIRVNGNADRRAPQETID 202  
 QY 124 CARDK-----NPIRIGVAGSLERDQKYEPTPOALLFESA 161  
 DB 203 YTEDEVQKELQHEQVFTPLVEKCKRYGAMRGTHGSLSDRIMSYTD-SPRGVESA 261  
 QY 162 MRVVDHLDRINFDQFVSVKASDVLAVESYRLANQI-----DQPLHIGTEAGARGSG 216  
 DB 262 FEFARICRKIDYHNFVFSKASNPVIMOVAYRLVAMVYHGWDPYHLGVTEAGEGEGD 321  
 QY 217 AVKSAIGLGLLSEGIQDTRVSLADPVPEEI 248  
 DB 322 RKMSAIGITLLQDGLDITIRVSLTEPPEEII 353

## RESULT 14

AAB45693 standard; Protein; 679 AA.

AAB45693;

15-MAR-2001 (first entry)

P. falciparum gcpe protein.

Isoprenoid; gcpe; yfgb; antimicrobial; transgenic plant; agriculture; antitumor; antiparasitic; antiviral; fungicidal; herbicidal.

Plasmodium falciparum.

MO200072022-A1.

30-NOV-2000.

20-MAY-2000; 2000MO-EP04592.

21-MAY-1999; 99DE-1023567.

21-MAY-1999; 99DE-1023568.

(JOMA/) JOMAA H.

WPI; 2001-025196/03.

N-PSDB; AAC82654.

Incorporating gcpe and yfgb genes into viruses and cells, for increasing isoprenoid content and identifying e.g. antimicrobial agents, comprises using DNA sequences from bacteria or parasites

Disclosure; Page 21-23; 36pp; German.

This invention describes a novel method for incorporating gcpe and yfgb genes into viruses and cells for increasing isoprenoid content and identifying e.g. antimicrobial agents, comprising using DNA sequences (I) from the gcpe or yfgb genes of bacteria or parasites or DNA sequences (II) which hybridize to the specified genes or encode a plastid protein with the same biological activity as those encoded by the genes. The invention also describes (1) plant cells containing (I) or (II); (2) transformed plant cells, and transgenic plants regenerated from them, that contain (I) or (II); (3) determining the enzymatic activity of a gcpe protein; or (4) screening compounds (A) that have antimicrobial, antiparasitic or antiviral activity in humans or animals or antiviral, antiparasitic, fungicidal or herbicidal activity in plants. (I) and (II)

are used: (i) to increase the isoprenoid levels in viruses and cells; (ii) for determining the enzymatic activity of gcpe and yfgp proteins; and (iii) to identify compounds that inhibit activity of gcpe, i.e. potential antibacterial, antitumor, antiparasitic or antiviral agents for use in humans or animals, or antiviral, antiparasitic, fungicidal or herbicidal agents for agriculture.

Sequence 679 AA:

Query Match 24.76% Score 460; DB 22; Length 679;  
Best Local Similarity 38.93% Pred. No. 2.3e-37;  
Matches 112; Conservative 57; Mismatches 81; Indels 38; Gaps 8;

9 RRRKSTR-IYVGNVPIGAPIAVOSMTNTRTDTVEATVNOIKALERVGADIVRSVPTMD 67  
Db KRLPFRVYIGNVAKIGNNKTAIQTMA SCDRNVECVYQIRKCKDGLADIVRLTVGCVQ 176  
Oy 68 AAEAF-----KLIRQVNVPLVADIHFDYRIALKVAEYGVCLRINPQIGN----- 114  
Db 177 EAQASVHIKEKLENNVNIPLVDIHFNPKITALMAADV-EKIRVNPNGVYDGRKKWIDK 235  
Oy 115 -----EERIRNVVDCARDKNIPIRIGVNASLSEKIDQEKYGEPTPOALL 158  
Db 236 VYKTRKEPDEGKLPIKEFVPLIECKR-LNRAIRIGTHGFLSSRYLSYGD-TPLALV 293  
Oy 159 ESAMRHVDHLRNLNDFQKVSVKASDVFLAVESYRL-AKOIDD-----PLHLGITTEAGGA 213  
Db 294 ESAFEFSDLCENNFNNLVFSKASNAVYMIQSTRLLVSKOYERNNMFPPIHLGVTGAGFG 353  
Oy 214 RSGAVKSAIGLLLSGIGDTLRVSLAADVEEIKVGFIDILKSLRIR 261  
Db 354 DNGRIKSYLGISLLYDGIPTIRISLDEPWEELTPCKKLVENLKR 401

# RESULT 15

AAW90883 standard; Protein: 752 AA.

AAW90883;

07-JUL-2000 (first entry)

P. falciptarum gcpe protein.

1-deoxy-D-xylulose-5-phosphate reductoisomerase; isoprenoid biosynthesis;

deoxy-D-xylulose-5-phosphate 1-deoxy-D-xylulose-5-phosphate synthase;

gcpe; antimicrobial; antibiotic; antiviral; antiparasitic; antimicrobial;

bactericide; fungicide; herbicide; treatment; disease.

Plasmodium falciptarum.

WO20017233-A2.

30-MAR-2000.

22-SEP-1999; 99WO-EP07035.

22-SEP-1998; 98DE-1043209.

21-MAY-1999; 99DE-1023567.

(JOMAA/) JOMAA H.

Jomaa H;

WPI: 2000-283543/24.

N-PSDB: AAW82996.

New nucleic acid, useful therapeutically and to screen for e.g.

antimicrobials and herbicides, encode proteins involved in isoprenoid

biosynthesis by the deoxy-D-xylulose-phosphate route

This invention describes novel polypeptides and their analogs which are involved in isoprenoid biosynthesis by the deoxy-D-xylulose-5-phosphate route. The proteins described include 1-deoxy-D-xylulose-5-phosphate reductoisomerase, 1-deoxy-D-xylulose-5-phosphate synthase and the gcpe protein. The products of the invention have antitumor, antibacterial, antiparasitic, antimicrobial, bactericide, fungicide and antiviral activity. The encoding nucleic acid sequences of the invention are used for prevention or treatment of diseases in humans and animals and also to screen compounds enzyme inhibitory activity (potential antimicrobials, antiparasitic agents, antivirals, fungicides, bactericides and herbicides, for use in human or veterinary medicine or agriculture). This sequence represents the Plasmodium falciptarum gcpe protein described in the method of the invention.

Sequence 752 AA:

Query Match 23.98% Score 446; DB 21; Length 752;  
Best Local Similarity 38.28% Pred. No. 7.1e-36;  
Matches 110; Conservative 56; Mismatches 84; Indels 38; Gaps 8;

9 RRRKSTR-IYVGNVPIGAPIAVOSMTNTRTDTVEATVNOIKALERVGADIVRSVPTMD 67  
Db KRLPFRVYIGNVAKIGNNKTAIQTMA SCDRNVECVYQIRKCKDGLADIVRLTVGCVQ 176  
Oy 68 AAEAF-----KLIRQVNVPLVADIHFDYRIALKVAEYGVCLRINPQIGN----- 114  
Db 177 EAQASVHIKEKLENNVNIPLVDIHFNPKITALMAADV-EKIRVNPNGVYDGRKKWIDK 235  
Oy 115 -----EERIRNVVDCARDKNIPIRIGVNASLSEKIDQEKYGEPTPOALL 158  
Db 236 VYKTRKEPDEGKLPIKEFVPLIECKR-LNRAIRIGTHGFLSSRYLSYGD-TPLGAV 293  
Oy 159 ESAMRHVDHLRNLNDFQKVSVKASDVFLAVESYRL-AKOIDD-----PLHLGITTEAGGA 213  
Db 294 ESAFEFSDLCENNFNNLVFSKASNAVYMIQSTRLLVSKOYERNNMFPPIHLGVTGAGFG 353  
Oy 214 RSGAVKSAIGLLLSGIGDTLRVSLAADVEEIKVGFIDILKSLRIR 261  
Db 354 DNGRIKSYLGISLLYDGIPTIRISLDEPWEELTPCKKLVENLKR 401

Search completed: August 12, 2003, 09:56:03  
Job time : 88 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2003, 09:54:00 ; Search time 30 Seconds  
(Without alignments)  
524.655 Million cell updates/sec

Title: US-09-921-992-78

Perfect score: 1866

Sequence: 1 MHQAPLQRRKSTRIVGVN.....RAKASQDEARIDVQVEK 372

Scoring table: BLOSUM62

Searched: Gap0 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PCtus.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/Backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1800	96.5	384	2	US-08-827-190-6 Sequence 6, Appl
2	1800	96.5	384	2	US-09-170-187-6 Sequence 6, Appl
3	1632.5	87.0	365	2	US-08-827-190-5 Sequence 5, Appl
4	1632.5	87.0	365	4	US-09-170-187-5 Sequence 5, Appl
5	1372.5	73.6	547	4	US-09-252-991A-22991 Sequence 22991, A
6	1195	64.0	378	4	US-09-328-352-7906 Sequence 7906, Ap
7	470.5	25.2	621	4	US-09-198-452A-389 Sequence 389, App
8	112.5	6.0	1058	4	US-09-252-991A-29105 Sequence 29105, A
9	111.5	6.0	578	3	US-09-066-046-6 Sequence 6, Appl
10	106.5	5.7	980	4	US-09-252-991A-30838 Sequence 30838, A
11	105.5	5.7	578	3	US-08-975-762-50 Sequence 50, Appl
12	105.5	5.7	578	3	US-09-295-028-50 Sequence 50, Appl
13	105.5	5.7	578	3	US-09-295-028-50 Sequence 50, Appl
14	101	5.4	465	4	US-09-106-582-50 Sequence 50, Appl
15	101	5.4	523	4	US-09-252-991A-29387 Sequence 29387, A
16	100.5	5.4	620	4	US-09-612-964-2 Sequence 2, Appl
17	100	5.4	343	3	US-09-328-352-7730 Sequence 7730, Ap
18	99.5	5.3	1374	4	US-09-039-859-2 Sequence 2, Appl
19	97	5.2	1037	4	US-09-252-991A-24636 Sequence 24636, A
20	96.5	5.2	705	4	US-08-328-352-7436 Sequence 4794, Ap
21	96.5	5.2	851	4	US-09-252-991A-24773 Sequence 7436, Ap
22	95.5	5.1	303	4	US-09-134-001C-4855 Sequence 24773, A
23	95.5	5.1	740	5	US-08-309-512-10 Sequence 4855, Ap
24	95.5	5.1	740	5	PCT-US92-08756A-10 Sequence 10, Appl
25	95	5.1	796	4	US-09-252-991A-17763 Sequence 17763, A
26	94.5	5.1	483	4	US-08-867-534A-51 Sequence 51, Appl
27	94.5	5.1	483	4	US-09-527-431-51 Sequence 51, Appl

28	94.5	5.1	659	4	US-09-252-991A-17904 Sequence 17904, A
29	94.5	5.1	1503	3	US-08-976-255-14 Sequence 14, Appl
30	93.5	5.0	619	4	US-09-252-991A-17411 Sequence 17411, A
31	93.5	5.0	659	4	US-09-252-991A-17731 Sequence 17731, A
32	93	5.0	459	4	US-09-491-785-2 Sequence 2, Appl
33	93	5.0	483	4	US-09-252-991A-27988 Sequence 27988, A
34	93	5.0	607	4	US-09-252-991A-20596 Sequence 20596, A
35	92.5	5.0	489	4	US-09-252-991A-27651 Sequence 27651, A
36	92.5	5.0	940	4	US-09-512-230C-2 Sequence 2, Appl
37	91	4.9	470	4	US-09-252-991A-26312 Sequence 26312, A
38	91	4.9	817	4	US-09-252-991A-31147 Sequence 31147, A
39	91	4.9	1181	4	US-09-252-991A-18480 Sequence 18480, A
40	90.5	4.8	484	4	US-09-252-991A-20787 Sequence 20787, A
41	90.5	4.8	489	4	US-09-252-991A-19810 Sequence 19810, A
42	90.5	4.8	573	4	US-09-252-991A-26428 Sequence 26428, A
43	90.5	4.8	1612	1	US-08-169-927-2 Sequence 2, Appl
44	90	4.8	358	4	US-09-784-508-4 Sequence 4, Appl
45	90	4.8	629	4	US-09-252-991A-31575 Sequence 31575, A

## ALIGNMENTS

```
RESULT 1
US-08-827-190-6
: Sequence 6, Application US/08827190
: Patent No. 5858367
:
: GENERAL INFORMATION:
: APPLICANT: Rather, Philip N.
: TITLE OF INVENTION: Methods For Screening For Antimicrobials
: TITLE OF INVENTION: Utilizing aarc And Compositions Thereof
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESS: Medlen & Carroll, LLP
: STREET: 220 Montgomery Street, Suite 2200
: CITY: San Francisco
: STATE: California
: COUNTRY: United States of America
: ZIP: 94104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/827,190
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Carroll, Peter G.
: REGISTRATION NUMBER: 32,837
: REFERENCE/DOCKET NUMBER: CASE-02443
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 705-8410
: TELEFAX: (415) 397-8338
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 384 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
:
: US-08-827-190-6
:
: Query Match 96.5%; Score 1800; DB 2; Length 384;
: Best Local Similarity 96.9%; Pred. No. 4.9e-188;
: Matches 372; Conservative 0; Mismatches 0; Indels 12; Gaps 6;
:
: QY 1 MHQAPLQRRKSTRIVGVNPIGDGAPINAVOSMTNTTIVEATV--NQKALERVGADI 58
: DB 1 MHQAPLQRRKSTRIVGVNPIGDGAPINAVOSMTNTTIVEATVFNQKALERVGADI 60
: QY 59 VRVSVPTMDAAEFKLRKQGVNPLVADIHFDYRIALKVAEGVDC--LAINQNGNNE 116
```

Db 61 VRSVPTMDAAEAFRLIKQVNVPLVADIHEDYRIALKVAEYVDC--LRINPGNIGEE 120  
QY 117 RIRWVDCARDKNIPIRIGVANGSLKDLQEKYGEPTQALLLESAMR--HVDHLDRLNFD 174  
Db 121 RIRWVDCARDKNIPIRIGVANGSLKDLQEKYGEPTQALLLESAMRTHVDHLDRLNFD 180  
QY 175 QFKVSVKASDVFLAVESYRLAKQIDQPLHIGITEAGARGSGAVKSAI--GIGLLISGCI 232  
Db 181 QFKVSVKASDVFLAVESYRLAKQIDQPLHIGITEAGARGSGAVKSAI--GIGLLISGCI 240  
QY 233 GDTLRSVLAADPVEEIKVGFIDLKSLRIRSGINFIACPTCSROEFVY--GTVALROR 290  
Db 241 GDTLRSVLAADPVEEIKVGFIDLKSLRIRSGINFIACPTCSROEFVY--GTVALROR 300  
QY 291 LEDITPMDSIIGCVNPGFALVSTLGVGNGKSGLYEDGVKRDLD--NNDMIDOL 348  
Db 301 LEDITPMDSIIGCVNPGFALVSTLGVGNGKSGLYEDGVKRDLDFTNNDMIDOL 360  
QY 349 EARIKAKASQDEARRIDVOQYER 372  
Db 361 EARIKAKASQDEARRIDVOQYER 384

## RESULT 2

US-09-170-187-6  
Sequence 6, Application US/09170187  
Patent No. 6383745

GENERAL INFORMATION:  
APPLICANT: Rather, Phillip N.  
TITLE OF INVENTION: Methods For Screening For Antimicrobials  
TITLE OF INVENTION: Utilizing aarc And Compositions Thereof  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medien & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/170,187  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/827,190  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Peter G.  
REGISTRATION NUMBER: 32,837  
REFERENCE/DOCKET NUMBER: CASE-02443  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 384 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-170-187-6

Query Match 96.5%; Score 1800; DB 4; Length 384;  
Best local similarity 96.9%; Pred. No. 4.9e-188;  
Matches 372; Conservative 0; Mismatches 0; Indels 12; Gaps 6;

QY 1 MHNQAPIQRKSTRIVGVNPIGDGAPIVQSMNTNRTTVDVATV--NQIKALERVGADI 58

Db 1 MHNQAPIQRKSTRIVGVNPIGDGAPIVQSMNTNRTTVDVATV--NQIKALERVGADI 60  
QY 59 VRSVPTMDAAEAFRLIKQVNVPLVADIHEDYRIALKVAEYVDC--LRINPGNIGEE 116  
Db 61 VRSVPTMDAAEAFRLIKQVNVPLVADIHEDYRIALKVAEYVDC--LRINPGNIGEE 120  
QY 117 RIRWVDCARDKNIPIRIGVANGSLKDLQEKYGEPTQALLLESAMR--HVDHLDRLNFD 174  
Db 121 RIRWVDCARDKNIPIRIGVANGSLKDLQEKYGEPTQALLLESAMRTHVDHLDRLNFD 180  
QY 175 QFKVSVKASDVFLAVESYRLAKQIDQPLHIGITEAGARGSGAVKSAI--GIGLLISGCI 232  
Db 181 QFKVSVKASDVFLAVESYRLAKQIDQPLHIGITEAGARGSGAVKSAI--GIGLLISGCI 240  
QY 233 GDTLRSVLAADPVEEIKVGFIDLKSLRIRSGINFIACPTCSROEFVY--GTVALROR 290  
Db 241 GDTLRSVLAADPVEEIKVGFIDLKSLRIRSGINFIACPTCSROEFVY--GTVALROR 300  
QY 291 LEDITPMDSIIGCVNPGFALVSTLGVGNGKSGLYEDGVKRDLD--NNDMIDOL 348  
Db 301 LEDITPMDSIIGCVNPGFALVSTLGVGNGKSGLYEDGVKRDLDFTNNDMIDOL 360  
QY 349 EARIKAKASQDEARRIDVOQYER 372  
Db 361 EARIKAKASQDEARRIDVOQYER 384

## RESULT 3

US-08-827-190-5  
Sequence 5, Application US/08827190  
Patent No. 5858367

GENERAL INFORMATION:  
APPLICANT: Rather, Phillip N.  
TITLE OF INVENTION: Methods For Screening For Antimicrobials  
TITLE OF INVENTION: Utilizing aarc And Compositions Thereof  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medien & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/827,190  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Peter G.  
REGISTRATION NUMBER: 32,837  
REFERENCE/DOCKET NUMBER: CASE-02443  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 365 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-827-190-5

Query Match 87.0%; Score 1622.5; DB 2; Length 365;  
Best local similarity 88.4%; Pred. No. 1.2e-168;  
Matches 320; Conservative 24; Mismatches 17; Indels 1; Gaps 1;

QY 1 MHNQAPIQRKSTRIVGVNPIGDGAPIVQSMNTNRTTVDVATV--NQIKALERVGADI 60

```

Db 1 MHNESPKRRKSTRIVGNVPIGDGAPIAVQSMNTRTTDEATVQIQSLERGVDIYR 60
QY 61 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAEGVCLIRNPNGINERIR 120
Db 61 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAEGVCLIRNPNGINERIR 120
QY 121 VVDCARDKNIPRIGVNGSLEKIDQEKYGEPTPOALLSARHVDLDRNFDOFKVSV 180
Db 121 VVDSARHNPRIPIRIGVNGSLEKIDQEKYGEPTPEALVESAMRHVDLDRNFDOFKVSV 180
QY 181 KASDVFLAVESYRLAKQIDOPHLGITEGAGRSAGVSAIGLLSEIGDITRLISL 240
Db 181 KASDVFLAVESYRLAKQIDOPHLGITEGAGRSAGVSAIGLLSEIGDITRLISL 240
QY 241 AADPVEIKVGFILKSLRIRSGINFACPTCSROEFYIGTVNLEORLEDIITPMYV 300
Db 241 AADPVEIKVGFILKSLRIRSGINFACPTCSROEFYIGTVNLEORLEDIITPMYV 300
QY 301 SIIGCVVNGPGEALVSTLGVTKGSKGLYEDGVR-KDRLDNDMDIDLEARIARAKASOL 359
Db 301 SIIGCVVNGPGEALVSTLGVTKGSKGLYEDGVR-KDRLDNDMDIDLEARIARAKASOL 359
QY 360 DE 361
Db 361 DE 362

```

#### RESULT 4

US-09-170-187-5  
Sequence 5, Application US/09170187  
Patent No. 6383745

GENERAL INFORMATION:  
APPLICANT: Rather, Philip N.  
TITLE OF INVENTION: Methods For Screening For Anticarbinals  
TITLE OF INVENTION: Utilizing aarc And Compositions thereof  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/170,187  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/827,190  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Peter G.  
REGISTRATION NUMBER: 32,837  
REFERENCE/DOCKET NUMBER: CASE-02443  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 365 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-170-187-5

Query Match 87.0%; Score 1622.5; DB 4; Length 365;  
Best Local Similarity 88.4%; Pred. No. 1.2e-168;

```

Matches 320; Conservative 24; Mismatches 17; Indels 1; Gaps 1;
QY 1 MHNOAPLQRRKSTRIVGNVPIGDGAPIAVQSMNTRTTDEATVQIQSLERGVADIYR 60
Db 1 MHNESPKRRKSTRIVGNVPIGDGAPIAVQSMNTRTTDEATVQIQSLERGVADIYR 60
QY 61 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAEGVCLIRNPNGINERIR 120
Db 61 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAEGVCLIRNPNGINERIR 120
QY 121 VVDCARDKNIPRIGVNGSLEKIDQEKYGEPTPOALLSARHVDLDRNFDOFKVSV 180
Db 121 VVDSARHNPRIPIRIGVNGSLEKIDQEKYGEPTPEALVESAMRHVDLDRNFDOFKVSV 180
QY 181 KASDVFLAVESYRLAKQIDOPHLGITEGAGRSAGVSAIGLLSEIGDITRLISL 240
Db 181 KASDVFLAVESYRLAKQIDOPHLGITEGAGRSAGVSAIGLLSEIGDITRLISL 240
QY 241 AADPVEIKVGFILKSLRIRSGINFACPTCSROEFYIGTVNLEORLEDIITPMYV 300
Db 241 AADPVEIKVGFILKSLRIRSGINFACPTCSROEFYIGTVNLEORLEDIITPMYV 300
QY 301 SIIGCVVNGPGEALVSTLGVTKGSKGLYEDGVR-KDRLDNDMDIDLEARIARAKASOL 359
Db 301 SIIGCVVNGPGEALVSTLGVTKGSKGLYEDGVR-KDRLDNDMDIDLEARIARAKASOL 359
QY 360 DE 361
Db 361 DE 362

```

#### RESULT 5

US-09-252-991A-22991  
Sequence 22991, Application US/09252991A  
Patent No. 6551795

GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT FILING DATE: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO: 22991  
LENGTH: 547  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-22991

Query Match 73.6%; Score 1372.5; DB 4; Length 547;  
Best Local Similarity 73.8%; Pred. No. 4.9e-141;  
Matches 267; Conservative 46; Mismatches 48; Indels 1; Gaps 1;

```

QY 1 MHNOAPLQRRKSTRIVGNVPIGDGAPIAVQSMNTRTTDEATVQIQSLERGVADIYR 60
Db 179 IHSASPTLRKSRKRWGVNVPVGDAPIAVQSMNTETCDVAATQAJRRLEDAADIVR 238
QY 61 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAEGVCLIRNPNGINERIR 120
Db 239 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAEGVCLIRNPNGINERIR 120
QY 121 VVDCARDKNIPRIGVNGSLEKIDQEKYGEPTPOALLSARHVDLDRNFDOFKVSV 180
Db 121 VVDSARHNPRIPIRIGVNGSLEKIDQEKYGEPTPEALVESAMRHVDLDRNFDOFKVSV 180
QY 181 KASDVFLAVESYRLAKQIDOPHLGITEGAGRSAGVSAIGLLSEIGDITRLISL 240
Db 181 KASDVFLAVESYRLAKQIDOPHLGITEGAGRSAGVSAIGLLSEIGDITRLISL 240
QY 359 KASDVFLAVESYRLAKQIDOPHLGITEGAGRSAGVSAIGLLSEIGDITRLISL 359

```





```

QY 134 IGVN-----AGSLEKDLQEKYGEPTPOLLESANRHYDHLRLNDFQKYSKASDVELA 188
D 414 --INHALLSGTTHIRPLSDAQFEGPRISVE-----ILKRTTPEFLV 453
QY 189 VESRLAKQIDOPHLGITEBAGARSGAKVSAIGLGLLSEGID-----234
D 454 DEIDGLKSAIDS---SVIE---LRLGNVGDNSPLRLRNFHVYKSHIPGIDRSLSYRV 506
QY 235 -----TLKYSLAAD-----PYEIKVGFIDILKSLRINSRGINFACIPCSQOE 278
D 507 NMALMYCTLSVSQLMNQERIEFWPIKE-TTMSDADDTITQALGI-----TRETLEY 559
QY 279 DVIGTVNABORLEDTITPMDVSIIGCVN---GPGCALVSTLGVGNGKSGLYEDGVR 335
D 560 LFSNDLQOLPHNLQ-----SLSRNLAMWRSNDGHRLETLALITGLNLADLSSEGR 611
QY 336 KDRLDNNDMDQLEARRAKAS 357
D 612 PERLEHD-----RIRLAKAS 626

```

## RESULT 9

```

; Sequence 6, Application US/0906046A
; Patent No. 6204252

```

## GENERAL INFORMATION:

```

APPLICANT: MORPHY, Cheryl
            STOREY, James
            BELTZ, Gerald A.
            COUGHLIN, Richard T.
TITLE OF INVENTION: CHARACTERIZATION OF GRANTLOCYTIC
                    EHRLICHIA AND METHODS OF USE
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESS: HALE AND DORR LLP
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: United States
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/066,046A
FILING DATE: 24-Apr-1998
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Superko, Colleen
REGISTRATION NUMBER: 39,850
REFERENCE/DOCKET NUMBER: 106,941,155
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 526-6000
TELEFAX: (617) 526-5000
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 578 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```

```

US-09-066-046-6

```

```

Query Match      6.0%; Score 111.5; DB 3; Length 578;
Best Local Similarity 18.6%; Pred. No. 0.0039;
Matches 88; Conservative 77; Mismatches 172; Indels 137; Gaps 20;
QY 7 IQRRKSTRIVGNVP-----IGDGA-----PIAVGSMNTNRT-TDEATVANO- 47

```

```

D 112 VQEEGGMGLINAPKAVVRFFKIEKSAEEDQYVDPVSVEATSGVDYTOEEOIDQ 171
QY 48 IKALERGADIVSVPTMDAAEAFKIKQVNVPIYADHFDYRIALKAVEGVCLRT 107
D 172 APAIEVEETEEQEV---ILEGTLIDQVAVQVPAVEELP---GVEAAEIVPSL-- 223
QY 108 NPGNIGNEERIRNV-----DCARDKNIPR-----IGNAGSLEKDLQEKYGEPT 154
D 224 -----EENKIQEVVVAPEAQLESAPESAPAPQESTVLGVAEGLKSEVSVEANADVP 277
QY 155 QALLSARHYDHLRLNDFQKYSV-----180
D 278 QKEVIGSQOQOEIAEALBEGTEADPEVEKETEVLKEDTLIDQPAQVPAEALPGV 337
QY 181 -----KASDVLAESYRL-LAKQIDOPHL-----LIGTEGARGSGAVKA 221
D 338 EAAEAIVPSLEENKIQEVVVAPEAQLESAPESAPQESTVLGTE-----GDKSE 391
QY 222 IGL-----GLLSRGIDTLRVSILADPVEIKYGF-----IKSLRINSRGINFIA 269
D 392 VSVEADAGMOQEAIGSD--QETQATEVEKEVEVSETEBEVEVILEEGTLIDQPAQ 449
QY 270 CPTCSROEPDVIGTVNAL-----EORLEDTITPMDVSIIGCV--VNGPGCALVSTLGVG 322
D 450 VPVVAEALPGVEAAAIYPSLEENKIQEVVVAPEAQLESAPESAPVQPESTVLGTE 509
QY 323 GNKSGSL-----YEDGVRRKDRIDNNDMDQLEARRAKASQLEARRIDV 367
D 510 GDLKSEVSVEADAGMOQEAIGSDQETQATEVEKEVEVSEADAGMOQEL--VDV 561

```

## RESULT 10

```

; US-09-252-991A-30838
; Sequence 30838, Application US/09252991A
; Patent No. 6551795

```

## GENERAL INFORMATION:

```

APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 30838
LENGTH: 980
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30838

```

```

Query Match      5.7%; Score 106.5; DB 4; Length 980;
Best Local Similarity 20.7%; Pred. No. 0.003;
Matches 88; Conservative 65; Mismatches 159; Indels 113; Gaps 20;

```

```

QY 16 YGVNPIGDGAPIAVGSMTNRTTDEATVANOIKALERVAGDIVASVPTMDAAEAFRLI 75
D 434 FVDVYRLPPGAPRSTRIRIVADVVG---EGRHPPVPAMTAPDLOANSLGV- 487
QY 76 KQVNVPLVADH---DYRI-----ALKVAEYGVDCRLINPGNIGNEERIRIMVYDC 124
D 488 --QVGIAVGVVEFEYGRRLRBCRAVVDQEAQVNAQ-----RVAPGGAAGVHAELLVYV 539
QY 125 ARDKNIPRIGVNGSLEKDLQEKYGEPT---PQALLS-----ANRHYDHLRLNFD 174
D 540 TAAQOLPT-----LATVLAEVYGEPAFVQPRVDAGFAADLVLLPVAAEQGVMAFG 591
QY 175 QFKYSKASDVFLAVESRLAKQIDOPHLGI-----TEGARGSG 216
D 592 QGQVYLPQOAVATGIEA-----RYETAILIGILVLAAPGLGHQPIERSEAGTA-AD 644

```

QY 217 AVKSAIGLGLISGIDTILRVSLADPVEIKVGFPIILKSRIRSGINFACPTCSRQ 276  
 Db 645 QVLLAVGVG--IERGVENLIVVRLCPERQAEFPAL-----VEAIAAQAQ 693  
 QY 277 EFDVIT-----GTVALEORLEDITTPMDVSIIGCVVNGPGEALVSTLG--VTGNNK 326  
 Db 694 VAAVAVSWLTKRSKGAIGIDPRRIE--AAHPAAVAVGAADPGALOGIMGAVTGGEE 752  
 QY 327 SGIV-----EEGVKDR-----LDNNDMIDQLEARI--RAKASOLDDEARRIDV 367  
 Db 753 LGALVAAAPGEDLDHPADGLRAVQAGTRPADHLDALQHLRQVLEGRAS-----AGRADL 808  
 QY 368 QOVER 372  
 Db 809 DAVDQ 813

## RESULT 11

US-08-975-762-50  
 Sequence 50, Application US/08975762  
 Patent No. 6207169

GENERAL INFORMATION:  
 APPLICANT: Reed, Steven G.  
 APPLICANT: Lodes, Michael J.  
 APPLICANT: Houghton, Raymond  
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND  
 NUMBER OF SEQUENCES: 73  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: SEED and BERRY LLP  
 STREET: 6300 Columbia Center, 701 Fifth Avenue  
 CITY: Seattle  
 STATE: Washington  
 COUNTRY: USA  
 ZIP: 98104

## TREATMENT

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/975,762  
 FILING DATE: 21-MAR-1997  
 CLASSIFICATION: 424  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Maki, David J.  
 REGISTRATION NUMBER: 30,392  
 REFERENCE/DOCKET NUMBER: 210121.439  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 206-622-4900  
 TELEFAX: 206-682-6031  
 INFORMATION FOR SEQ ID NO: 50:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 578 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-975-762-50

Query Match 5.7%; Score 105.5; DB 3; Length 578;  
 Best Local Similarity 19.5%; Pred. No. 0.018;  
 Matches 91; Conservative 77; Mismatches 176; Indels 123; Gaps 21;

QY 7 IQRRKSTRIVGNVP-----IGDGA-----PIAVOSMTNRT--TVEATVNO- 47  
 Db 112 VQEEETGMVLLINAPKAVVRFKLEKSAEPPQYVDSVVSATGSGVDTQEEQIDOE 171  
 QY 48 IKALERVGAD-----IVRSVPTMDAAEAF--KLIKQVNV 82  
 Db 172 APAIEVEETEEOEVLIEBGTLLIDLPQVAVPVVAEALPGVEAAEAIYPSLENNKLOEV 231  
 QY 83 LVADHFDYRIALKV--AEYGVDCLRINPGNIGNEERIRNVVDCARDKNIPIRIGVNA 139

Db 232 VVAPEAQLESAPESVAPAPPESTVLGVAAGDLKSEVSVEANDVAOKE-----VISG 284  
 QY 140 SLEKDLOEK--GEPTQALLE--SAMRHVDHRLNFDQKVS----- 180  
 Db 285 QQBOELAEALGEAEPAVEVEETEVILKEETLLIDLPQVAVPVVAEALPGVEAAEAIY 344  
 QY 181 -----KASDVLAVESYRL-LAKQIDOPH-----LGITEAGARSQAVKSAIGL---- 224  
 Db 345 PSLENNKLOEVVVAPEAQLESAPESVAPAPPESTVLGVT-----GDLKSEVSVEADA 398  
 QY 225 GLLISEGIDTLRVSLADPVEIKVGF-----ILSKLIRSGINFACPTCSRQ 276  
 Db 399 GMOQEGAGISD--QETQATEEVEKVESVETKTEPEVILLEGTLIDLPQVAVPVVAEA 456  
 QY 277 EFDVITGVNAL-----EORLEDITTPMDVSIIGCV--VNGPGEALVSTLG--VTGNNKSGL 329  
 Db 457 ELPGVEAAEAIYPSLENNKLOEVVVAPEAQLESAPESVAPAPPESTVLGVTGDLKSEV 516  
 QY 330 -----YEDGVKDRDLNNDMIDQLEARIKAKASOLDDEARRIDV 367  
 Db 517 SVADAGMOQEGAGISDQETQATEEVEKVESVADAGMOQL--VDV 561

## RESULT 12

US-09-295-028-50  
 Sequence 50, Application US/09295028  
 Patent No. 6277381

GENERAL INFORMATION:  
 APPLICANT: Reed, Steven G.  
 APPLICANT: Lodes, Michael J.  
 APPLICANT: Houghton, Raymond L.  
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
 TITLE OF INVENTION: AND TREATMENT OF EHRLICHIA INFECTION  
 FILE REFERENCE: 210121.439C4  
 CURRENT APPLICATION NUMBER: US/09/295,028  
 CURRENT FILING DATE: 1999-04-20  
 NUMBER OF SEQ ID NOS: 85  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 50  
 LENGTH: 578  
 TYPE: PROT  
 ORGANISM: Ehrlichia sp.  
 US-09-295-028-50

Query Match 5.7%; Score 105.5; DB 3; Length 578;  
 Best Local Similarity 19.5%; Pred. No. 0.018;  
 Matches 91; Conservative 77; Mismatches 176; Indels 123; Gaps 21;

QY 7 IQRRKSTRIVGNVP-----IGDGA-----PIAVOSMTNRT--TVEATVNO- 47  
 Db 112 VQEEETGMVLLINAPKAVVRFKLEKSAEPPQYVDSVVSATGSGVDTQEEQIDOE 171  
 QY 48 IKALERVGAD-----IVRSVPTMDAAEAF--KLIKQVNV 82  
 Db 172 APAIEVEETEEOEVLIEBGTLLIDLPQVAVPVVAEALPGVEAAEAIYPSLENNKLOEV 231  
 QY 83 LVADHFDYRIALKV--AEYGVDCLRINPGNIGNEERIRNVVDCARDKNIPIRIGVNA 139  
 Db 232 VVAPEAQLESAPESVAPAPPESTVLGVAAGDLKSEVSVEANDVAOKE-----VISG 284  
 QY 140 SLEKDLOEK--GEPTQALLE--SAMRHVDHRLNFDQKVS----- 180  
 Db 285 QQBOELAEALGEAEPAVEVEETEVILKEETLLIDLPQVAVPVVAEALPGVEAAEAIY 344  
 QY 181 -----KASDVLAVESYRL-LAKQIDOPH-----LGITEAGARSQAVKSAIGL---- 224  
 Db 345 PSLENNKLOEVVVAPEAQLESAPESVAPAPPESTVLGVT-----GDLKSEVSVEADA 398  
 QY 225 GLLISEGIDTLRVSLADPVEIKVGF-----ILSKLIRSGINFACPTCSRQ 276  
 Db 399 GMOQEGAGISD--QETQATEEVEKVESVETKTEPEVILLEGTLIDLPQVAVPVVAEA 456

QY 277 EFDVIGTVAL-----EORLEDTITPMDSIIIGCV--VNGPGEALVSTLGTGNGKSGI 329  
 Db 457 ELPGEAAEAIYPSLEENKLEDEYVVAPEAQOLESAPESAPQPESTVLGTEEDLKEV 516  
 QY 330 -----YEDVKRDRLDNNMDIDLEARIKASOLDREARIDV 367  
 Db 517 SVADAGMOQOEGISDQTEQATEEVEKEVEVSVEADAGMOQEL--VDV 561

## RESULT 13

US-09-106-582-50  
 ; Sequence 50, Application US/09106582  
 ; Patent No. 6306402

GENERAL INFORMATION:  
 APPLICANT: Reed, Steven G.  
 APPLICANT: Lodes, Michael J.  
 APPLICANT: Houghton, Raymond  
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND  
 NUMBER OF SEQUENCES: 73  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: SEED AND BERRY LLP  
 STREET: 6300 Columbia Center, 701 Fifth Avenue  
 CITY: Seattle  
 STATE: Washington  
 COUNTRY: USA  
 ZIP: 98104

## THERAPY

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/106,582  
 FILING DATE: 29-JUN-1998  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Makl, David J.  
 REGISTRATION NUMBER: 31,392  
 REFERENCE/DOCKET NUMBER: 210121.439C2  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 206-622-4900  
 TELEFAX: 206-682-6031

INFORMATION FOR SEQ ID NO: 50:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 578 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-106-582-50

Query Match 5.7%; Score 105.5; DB 4; Length 578;  
 Best Local Similarity 19.5%; Pred. No. 0.018;  
 Matches 91; Conservative 77; Mismatches 176; Indels 123; Gaps 21;

QY 7 IORRSTRIVGNP-----IGDGA-----PIAYOSMTTRT--TOVEARVNG- 47  
 Db 112 VOEEEGTGLIAPKAVRFRFKIKSAEPEQYVDPSEVATSGVDTQEEDEID 171  
 QY 48 IAKLENVGAD-----YRVSVPTMDAAEAF--KLIKOQVNP 82  
 Db 172 APAIEVEETEEOGVLIEGTLIDLEQPVAVPVAEALPGVEAAAIYPSLEENKLEV 231  
 QY 83 LVADHIFDRILAKY---AEYGVDCIRINPGINGNEERIRMYVDCARDKNIDIRIGVNG 139  
 Db 232 VVAPEAQOLESAPESAPQPESTVLGVAEGDLKSEVSVEANADVAQKE-----VLSG 284  
 QY 140 SLEKDLQERY-GEPTQALLE--SAMRHVDHLDRLEDFQKVS----- 180  
 Db 285 OQEOEIAEALLEGTEAPVEKEETEVEVLKEDTLIDLEQPVAVPVAEALPGVEAAEAIY 344  
 QY 181 -----KASDVFLAVESYRL-LAKIDQPLH-----LGITTEGARGSAVKSAILG---- 224

Db 345 PSEENKLEQEVVVAPEAQOLESAPESAPQPESTVLGTE-----GDKSEVSVEAD 398  
 QY 225 GLLSEIGDTLNVSLAADPVEIKYGF-----ILKSLIRSGINFIACPTCSHQ 276  
 Db 399 GMOQOEGISD--QETQATEEVEKEVEVSERKTEPEVILEEGTLIDLEQPVAVPVA 456  
 QY 277 EFDVIGTVAL-----EORLEDTITPMDSIIIGCV--VNGPGEALVSTLGTGNGKSGI 329  
 Db 457 ELPGEAAEAIYPSLEENKLEDEYVVAPEAQOLESAPESAPQPESTVLGTEEDLKEV 516  
 QY 330 -----YEDVKRDRLDNNMDIDLEARIKASOLDREARIDV 367  
 Db 517 SVADAGMOQOEGISDQTEQATEEVEKEVEVSVEADAGMOQEL--VDV 561

## RESULT 14

US-09-252-991A-29387  
 ; Sequence 29387, Application US/09252991A  
 ; Patent No. 6551795

GENERAL INFORMATION:  
 APPLICANT: Marc J. Rubenfield et al.  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 FILE REFERENCE: 107196.136  
 CURRENT APPLICATION NUMBER: US/09/252,991A  
 PRIOR FILING DATE: 1999-02-18  
 PRIOR APPLICATION NUMBER: US 60/074,788  
 PRIOR FILING DATE: 1998-02-18  
 PRIOR APPLICATION NUMBER: US 60/094,190  
 PRIOR FILING DATE: 1998-07-27  
 NUMBER OF SEQ ID NOS: 33142  
 SEQ ID NO 29387

LENGTH: 465  
 TYPE: PRT  
 ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-29387

Query Match 5.4%; Score 101; DB 4; Length 465;  
 Best Local Similarity 21.8%; Pred. No. 0.038;  
 Matches 81; Conservative 43; Mismatches 122; Indels 126; Gaps 15;

QY 25 GAPIAV-OSMTNRTTVEATVNOIKALERVGADIVRVSVPTMDAAEAFKLKQOVNPL 83  
 Db 59 GLEIALAETIVALLDLE---EDRADHVRGDEL-----QDQAVVARGAVDQVCPAQ 108  
 QY 84 VADHIFDRILAKVAEYGVDCIRINPGINGNEERIRMYVDC----- 125  
 Db 109 LLE-----RLAMP-GDAGVDLLV--GVGRHEHAAVVAQLAHGIDVVGQGVLDFA 160  
 QY 126 -----RDKNIPR-----IGVAGSLEKLDQEKGEPTQALLES 161  
 Db 161 MVFADEFLDNGIVGRFYVDRDADLPARGCHGAGDAGLADVEYANLAEGDLVETGP 220  
 QY 162 -----MRHVHLDRL-----NEDQKVSVAKSDVFLAVESYRLAKOIDQPLH 204  
 Db 221 DVHLAALDVQGVYVGGADRVVVGAAAFDELEVDAVVAVAVDEIQRTAQAQAD 280  
 QY 205 LGITPAGGA--RSQAVKSAIGL---LDSGIDTLRVSLAADPVEIKYGFILKSLR 259  
 Db 281 VQIAEVGVAHAQLGLTGVGVGGSVLPBPGHGAAGAAVLEGE----- 324  
 QY 260 IRSRGINFIACPTCSROEFVDTIGTVNALEQRLEDTITPMDSIIIGCVVNGPGEA----- 313  
 Db 325 -----FVDMRGRRAVGHV-----DIVLLEQEDFLIGTLMGPRGEAHVH 365  
 QY 314 --LVSTLGYTGG 323  
 Db 366 AQLDALGVGG 377

RESULT 15  
 US-09-612-964-2







```

1  APPLICANT: OCHIAI, KEIKO
2  APPLICANT: YOKOI, HARUHIKO
3  APPLICANT: TATEISHI, NAOKO
4  APPLICANT: SENOH, AKIHIRO
5  APPLICANT: IKEDA, MASATO
6  APPLICANT: OZAKI, AKIO
7  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
8  FILE REFERENCE: 249-125
9  CURRENT APPLICATION NUMBER: US/09/738, 626
10 CURRENT FILING DATE: 2000-12-18
11 PRIOR APPLICATION NUMBER: JP 99/577484
12 PRIOR FILING DATE: 1999-12-16
13 PRIOR APPLICATION NUMBER: JP 00/159162
14 PRIOR FILING DATE: 2000-04-07
15 PRIOR APPLICATION NUMBER: JP 00/280988
16 PRIOR FILING DATE: 2000-08-03
17 NUMBER OF SEQ ID NOS: 7059
18 SOFTWARE: PatentIn ver. 3.0
19 SEQ ID NO 5705
20
21 LENGTH: 378
22
23 TYPE: PRT
24 ORGANISM: Corynebacterium glutamicum
25 US-09-738-626-5705

```

Query Match	40.8%	Score 761	DB 10	Length 378
Best Local Similarity	44.3%	Pred. No. 1.3e-69		
Matches 159	Conservative 64	Mismatches 124	Indels 12	Gaps 3
QY	9	RRKSTRIVGVNPIJGDGAPIVQSKTNRTDVEATVQNIKALEVGDIVVSVPTDA	68	
Db	5	RRKTRQLMVAKGVGSDHPISVQSKTTKTHDINTDQIQLATGCDIVYVACPXYD	64	
QY	69	AEAEKLLKQOVNPLVADIHFDYRIALKVAEYGVDCERIINPNIGN-EERRIMVYDCARD	127	
Db	65	AEALPIIAKKSPIPIYADIHFPKPIFEAIDAGCAAVRVNPNIGKEFGRVKEVAKAAGD	124	
QY	128	KNIPTRIGVNGNSLEKIDQEKY-GEPTFOALLSEMRVHDHIDRLNFPQFKSVYASDYF	186	
Db	125	AGITRIGVNGNSLDKRLIDKRIHGAATPEALVESAMWAGLFEHGFDDIALSVYASHDPV	184	
QY	187	LAVESYRLAKOIDQPLHGLITTEAGGARSAGVASKAITGILLSEIGDILRVSLADPYE	246	
Db	185	LMVEATYRLAQSDYPLHLGYTVEAGPKRMGITSKSVAGCALLSSGIGTIRKLSADPYE	244	
QY	247	EIKVGFDLKSLRIRSRGINFIACPTCSRGQEDVYGTVALNEQRLEDITTPMDVSIICV	306	
Db	245	EIKVGDDILSLNLRPKLEIVSPSCGRADVDYSLAEVETALDGMELVRVAVMGCV	304	
QY	307	VNGVGEALVSTLGYTGKMKSGLYEDVGRKRRLDNNDMMIDOLEAIRAKASOLDARRI	365	
Db	305	VNGGEARDADLVGASNGKQILEVKGVIKTVSPESQIVELT-----IEEAMRI	353	

RESULT 5  
 US-09-712-363-256  
 : Sequence 256 Application US/09712363  
 : Patient No. US20020164588A1  
 :  
 : GENERAL INFORMATION:  
 :  
 : APPLICANT: Eisenberg, David  
 : APPLICANT: Rotstein, Sergio H.  
 : APPLICANT: Marcotte, Edward M.  
 : TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND  
 : FILE REFERENCE: 07419-032001  
 :  
 : CURRENT APPLICATION NUMBER: US/09/712,363  
 :  
 : PRIOR APPLICATION NUMBER: 2000-11-13  
 : PRIOR FILING DATE: 2000-01-28  
 :  
 : PRIOR APPLICATION NUMBER: 60/179,531  
 : PRIOR FILING DATE: 2000-02-01  
 : PRIOR APPLICATION NUMBER: 60/117,844  
 : PRIOR FILING DATE: 1999-01-29  
 : PRIOR APPLICATION NUMBER: 60/118,206,

```

? PRIOR FILING DATE: 1999-02-01
? PRIOR APPLICATION NUMBER: 60/126,593
? PRIOR FILING DATE: 1999-03-26
? PRIOR APPLICATION NUMBER: 60/134,093
? PRIOR FILING DATE: 1999-03-14
? PRIOR APPLICATION NUMBER: 60/134,092
? PRIOR FILING DATE: 1999-05-14
? PRIOR APPLICATION NUMBER: 60/165,124
? PRIOR FILING DATE: 1999-11-12
? PRIOR APPLICATION NUMBER: 60/165,086
? PRIOR FILING DATE: 1999-11-12
? NUMBER OF SEQ ID NOS: 292
? SOFTWARE: PstSeq for Windows Version 4.0.
? SEQ ID NO: 256
? LENGTH: 387
? TYPE: PRT
? ORGANISM: Mycobacterium tuberculosis
IS-09-712-363-256

```

[illegible]

```

US-RESULT 6
US-10-156-761-10088
? Sequence 10088, Application US/10156761
? Publication No. US20030119018A1
? GENERAL INFORMATION:
? APPLICANT: OMURA, SATOSHI
? APPLICANT:
? APPLICANT: IKEDA, HARUO
? APPLICANT: HOSHIKAWA, JUN
? APPLICANT: HORIKAWA, HIROSHI
? APPLICANT: SHIBA, TADAYOSHI
? APPLICANT: SAKAKI, YOSHIYUKI
? APPLICANT: HAYTORI, MASAHIRA
? TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
? FILE REFERENCE: 249-262
? CURRENT APPLICATION NUMBER: US/10156, 761
? PRIORITY FILING DATE: 2002-05-29
? PRIOR APPLICATION NUMBER: JP 2001-204089
? PRIOR FILING DATE: 2001-05-30
? PRIOR APPLICATION NUMBER: JP 2001-272697
? PRIOR FILING DATE: 2001-08-02
? NUMBER OF SEQ ID NOS: 15109

```

SEQ ID NO 10088  
LENGTH: 1046  
TYPE: PRT  
ORGANISM: Streptomyces avermitilis  
US-10-156-761-10088

Query Match 5.8%; Score 108; DB 15; Length 1046;  
Best Local Similarity 21.8%; Pred. No. 0.094;  
Matches 86; Conservative 48; Mismatches 139; Indels 122; Gaps 18;

QY 11 KSTRIVGNVPIGGA-PIAVOSMTNRTTVEATNOIKALERVADIVRVSPTMDAA 69  
DB 614 KSTDAIIVVANGVMPQTEALNHAKADVPYIV-AVNRIDVAGAD-----PT----- 662  
QY 70 EAFKIKQOVNPIVADHFDYRIALAKYAEVGDCLRNPGNGISEERIRWVDCARDKN 129  
DB 663 ---KRGQTERGLVAE-----ETGGD-----TMRVD----- 686  
QY 130 IPIRIGNVAGSL-----EKDIQKYGEPPTQALLSARHVDHLRINFDQFV- 178  
DB 687 ISAKGLNIESLLEAVVLTAADASDLRANPEQDAGIAIES-----HLDGRGAVATVL 740  
QY 179 ----SVKSDVFLAVESRLAKQIDQFLHGITAGARGSAVSAIGLILLEGIGD 234  
DB 741 VQRGRLRVGDMVYGDAGRVARAMDDEKGE-NVEAGSPTPLY-----LGLTNVPGAGD 794  
QY 235 TLRVSLADPVEEIKVGFILKSLRIRSGINFI-----ACPTCSROEPDYI 281  
DB 795 NFLV-----VDETRAQIAEKRAKERNANFARGRVFSLNDELAKGLVQELNI 848  
QY 282 -----GTVMLEQRL-EDITTPMDSIIGCVNGPEALVSTLGVTKGNKSGLYEDG 333  
DB 849 IKGDASGSVEALESSLQDVGEEVDIRLH---RGVAGVTESDINLATG----- 895  
QY 334 VRKRLDNNMDQLEARIKAKASQLDERRIDVQ 368  
DB 896 -----SDATVIGFNVRAGRAAQAAREGVDR 923

## RESULT 7

US-09-159-469-50  
Sequence 50, Application US/09159469  
Patent No. US2002006453A1

GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Houghton, Raymond  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND  
TITILE OF INVENTION: THERAPY OF EHRlichia INFECTION  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA

ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/159,469  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/106,582  
FILING DATE: 29-JUN-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.439C2

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-622-4900  
TELEFAX: 206-682-6031  
INFORMATION FOR SEQ ID NO: 50:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 578 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-09-159-469-50

Query Match 5.7%; Score 105.5; DB 9; Length 578;  
Best Local Similarity 19.5%; Pred. No. 0.066;  
Matches 91; Conservative 77; Mismatches 176; Indels 123; Gaps 21;

QY 7 IQRKSTRIYGNVP-----IGDA-----PIAVOSMTNRT-TDVEATVNO- 47  
DB 112 VOEEGTMYLINAEKAVRFEKLSAEEPQTVDSVYESATGSGVDTQEEQIDOE 171  
QY 48 IKALERVAGAD-----IVRSVPTMDAAEAF-KLIKQOVNVP 82  
DB 172 APAIEVEETEDEEVILEEGTLIDEPVAAQVPVVAEALPVEAAEALVPSLEENKIQEV 231  
QY 83 LVADIHFDYRIALKY---AEVGDCLRNPGNGISEERIRWVDCARDKNIPIRIGNVAG 139  
DB 232 VVAPPAQOLESAPESAPQPESTVLGVAEGDLKSEVSSEANADVAQRE-----VLSG 284  
QY 140 SLEKDLQEKY-GEPTQALLE--SAMRHVDHLRINFDQFVY----- 180  
DB 285 QOEQELAELESTETPEVEKEETEVLKEDTLIDEPVAAQVPVVAEALPVEAAEALV 344  
QY 181 -----KASDVFLAVESYRL-LAKQIDQLH-----LGITAGARGSAVSAIGL- 224  
DB 345 PSLEENKIQEVVVAPEAQOLESAPESAPQPESTVLGVTGTE-----DLKSEVSVEADA 398  
QY 225 GILLSEGIQDITLRSVSLADPVEEIKVGF-----ILKSLRIRSGINFIACPTCSRO 276  
DB 399 GMOQEGAGISD-QETQATEEVEKEVSVETKTEEPEVILEEGTLIDEPVAAQVPVVAE 456  
QY 277 EFDVIGTVNAL-----EORLEDIITPMDSIIGCV--VNGPEALVSTLGVTKGNKSGL 329  
DB 457 ELPGVBAEALVPSLEENKIQEVVVAPEAQOLESAPESAPQPESTVLGVTGGLKSEV 516  
QY 330 -----YEDGVKRLDNNMDQLEARIKAKASQLDERRIDV 367  
DB 517 SVEADAGMOQEGAGISDQETQATEEVEKEVSVVEADAGMOQL--VDV 561

## RESULT 8

US-09-798-042-50  
Sequence 50, Application US/09798042  
Patent No. US20020068343A1

GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Houghton, Raymond L.  
APPLICANT: McNeill, Patricia D.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
TITILE OF INVENTION: AND TREATMENT OF EHRlichia INFECTION  
FILE REFERENCE: 210121.439C7  
CURRENT APPLICATION NUMBER: US/09/798,042  
CURRENT FILING DATE: 2001-03-02  
NUMBER OF SEQ ID NOS: 108  
SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 50  
LENGTH: 578  
TYPE: PRT  
ORGANISM: Ehrlichia sp.  
US-09-798-042-50

Query Match 5.7%; Score 105.5; DB 9; Length 578;  
Best Local Similarity 19.5%; Pred. No. 0.066;



Matches 91; Conservative 77; Mismatches 176; Indels 123; Gaps 21;

```

QY 7 IORRSTRIVGNP-----IGDGA-----PIAVOSTNTT-TRT-DVEAVYNO- 47
Db 112 VQEEBGTMTLINPAKAVVRFRKTEKSAEEOQYVDPVSVEATSGVDIOEQEOLDOE 171
QY 48 IKALERVAD-----IVRVSVPTMDAAEAF--KLIRKQVNV 82
Db 172 APAIEVEETEEOEYILEGTLIDLEQVPAQVPAEAELEPGVAAEIVLSLEENKIOEY 231
QY 83 LVADIHFDYRIALAKV---AAYGVDCLRINPNTIGNNEERIMVDDCARDKNIPRIGVNAG 139
Db 232 VVAPPAQQLSAEVSAPAPQESTVLGVAEGDLKSEVSEANADVAKO-----VISG- 284
QY 140 SLEKDLQERY-GEPTPOALLE--SAMRHVDHLRNDQFKVSY-----180
Db 285 QOQOEIALEALEGAEVAVPEVEEVEVLKEPTLIDLEQVPAQVPAEAELEPGVAAEIV 344
QY 181 -----KASDELAVESYRL-LAKQIDQPLH-----LGTIEAGARSAGVAKSAIGL- 224
Db 345 PSLEENKLOEVVVAPEAQLESAPAPQESTVLGYTE-----GDLKSEVSEADA 398
QY 225 GLLSGIGPTLVLSLAADPVEIKYGF-----ILKSLRIRSGINFACPTCSRO 276
Db 399 GMOQEGASID--OETQATEEVEVEVSVETKTEPEVILEGTLIDLEQVPAQVPAEAE 456
QY 277 EFDVIGTVNAL-----EQRLEDIITPMVSIIGCV--VNGPGEALVSTLGVTGNKKSGL 329
Db 457 ELPGVAAEALVLSLEENKLOEVVVAPEAQLESAPAPQESTVLGYTEGDLKSEV 516
QY 330 -----YEOGVRRDRLDNMDIOLKRIKAKSOLDDEARRITV 367
Db 517 SVADAGMOQEGASIDQETOATEEVEKVEVSEADAGMOQEL--VDV 561

```

RESULT 9  
US-09-815-242-5635  
Sequence 5635, Application US/09815242  
Patent No. US20020061569A1

GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO: 5635  
LENGTH: 2368  
TYPE: PRT  
ORGANISM: Staphylococcus aureus

US-09-815-242-5635

Query Match 5.3%; Score 99; DB 9; Length 2368;  
Best Local Similarity 20.8%; Pred. No. 2.9;  
Matches 89; Conservative 58; Mismatches 155; Indels 126; Gaps 19;

```

QY 3 NOAPIQ-RRKSTIYVGNPFIGG-----APLAVOSTNTTTRT-DVE--- 42
Db 1532 NOAKNTINDOSTDEYVDNA-VKEGKAKINAKTFSEYKKDALKIEAAYNAKTEADNSN 1590
QY 43 -ATVNOIKALERVADIVRVSVPTMDAAEAFKLIRKQVNVPLVADHIFDRIKLKVAEY 101
Db 1591 ASTRSSIEAEKQKALIELKQADQNVQNS---KODIEQIINDLD-----1633
QY 102 VDCLRINPNTIGNNEERIMVDDC---ARDKNIPRIGVNAGSLEKDLQERYGEPTPOALL 158
Db 1634 ---NINDYTIPTGKRESATTDLYAYADQKKNISADTJNATQEQOQAIKOVQNVOTL 1689
QY 159 ESAMRRVVDH-----LDRLNFD-----QFKSVASDVFLAV 189
Db 1690 EGINNVGDNDVDALTGKRAIDAQVATYKPAKNOYIDAKAEETKESIDQDLTA 1749
QY 190 ESTYRLA--KQIDQPLHGTENGARSAGVAKSAIGLISGIG-DTLRVSLADPYE 246
Db 1750 EKTEALAMIKQITDQAKQGITDA--TTTAEVEKAKQGLEAFDNIIDSTEKQKA---IE 1804
QY 247 EIKVFGDILKSLRIRSGINFACPTCSROEFDVIGTVNALQERLEDIITPMVSIIGCV 306
Db 1805 ELFTALDQIEA-----GNVADATTEKE-----AFTNALBDDISKATEDISDQT 1850
QY 307 VNGPGEALVSTLGVTGNKKSGLYEDGVRRDRLDNMDIOLKRIKAKSOLDDEARRITV 363
Db 1851 TN---AETATV-----KSALEQLKAKQININPVKKNALEAIR 1884
QY 364 RIDVOQVE 371.
Db 1885 EYVNAKQIE 1892

```

RESULT 10  
US-09-815-242-12389

Sequence 12389, Application US/09815242  
Patent No. US20020061569A1

GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 12389  
 LENGTH: 2368  
 TYPE: PR  
 ORGANISM: Staphylococcus aureus  
 US-09-815-242-12389

Query Match 5.39% Score 99; DB 9; Length 2368;  
 Best Local Similarity 20.89% Pred. No. 2.9;  
 Matches 89; Conservative 58; Mismatches 155; Indels 126; Gaps 19;

3 NOAPIO-RRKSTRIVYENPIGDS-----APIAVQSMTRTTTDE--- 42  
 1532 NQATNIDQSTDEYVNDIA--KECKAKINAVKTESEYKKDALAKTEAAYNAKYTEADNSN 1590  
 43 -ATVNOIKALERVADIVRVSPFMDAAEAFKLIKQOVNVPVADIHEDYRIALKVAEYG 101  
 1591 ASTSEIAEAKOKIAEKEQTQDQVNOATS---KDLEVOIHNDLD----- 1633  
 102 VDCIRINPGNIGNEERGGMVDC---ARDKNIPRIGVNASLEKDLQEKYGEPTPOL 158  
 1634 ----NINDYTIPTGKKSATTDLYAYADQKKNNISADTNATODEKQAIKQVDQVOTAL 1689  
 159 ESAMRHVDH-----LDRLNFD-----QEKVSKASDVFLAV 189  
 1690 ESINNGVDNGVDALGKKAIDAIGVATVFKKANQVIDAKAEETKESIDQSLTAE 1749  
 190 ESKYLLA--KQIDPIHGTTEAGARGSAIGLIGLISEGIG--DLEKVSIAADPVE 246  
 1750 EKTALAAIKQITDOAKQGITDA--TTTAEVEKAKAQGLFEDNIDISTEKOKA---IE 1804  
 247 EIKVGFILKSLIRSGINFIACPTGSRQEFVIGTVNALFORLEDITTPMDVSIIGCV 306  
 1805 ELEFALQDIEA-----GVNVDADATTEKE-----AFTNALMEDILSKATEDISDQT 1850  
 307 VNGCEALVSTLGTGKKKSGLIEDGVKRDLDNNMIDOLEA-RIR--AKASLDLAR 363  
 1851 TN---AEIATV-----KNSALEDKAKORINPVVKKNALEAIR 1884  
 364 RIDVOQVE 371  
 1885 EYVKNQIE 1892

RESULT 11  
 US-10-156-761-8947

Sequence 8947, Application US/10156761  
 Publication No. US20030119016A1

GENERAL INFORMATION:  
 APPLICANT: OMURA, SATOSHI  
 APPLICANT: IKEDA, HARUO  
 APPLICANT: ISHIKAWA, JUN  
 APPLICANT: HORIKAWA, HIROSHI  
 APPLICANT: SHIBA, TADAYOSHI  
 APPLICANT: SAKAKI, YOSHIYUKI  
 APPLICANT: HATTORI, MASAHISA  
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 FILE REFERENCE: 249-262  
 CURRENT APPLICATION NUMBER: US/10/156,761  
 CURRENT FILING DATE: 2002-05-29  
 PRIOR APPLICATION NUMBER: JP 2001-204089  
 PRIOR FILING DATE: 2001-05-30  
 PRIOR APPLICATION NUMBER: JP 2001-272697  
 PRIOR FILING DATE: 2001-08-02  
 NUMBER OF SEQ ID NOS: 15109  
 SEQ ID NO 8947  
 LENGTH: 365  
 TYPE: PR  
 ORGANISM: Streptomyces avermitilis  
 US-10-156-761-8947

Query Match 5.39% Score 98.5; DB 15; Length 365;  
 Best Local Similarity 20.93% Pred. No. 0.17;  
 Matches 77; Conservative 54; Mismatches 138; Indels 99; Gaps 19;

18 GNVPFGDAPLAVQSMTRTTTDEVA-----TVNOIKALERVAD----- 57  
 28 GSAPVGDGAVLHITGDCAGAGSVFARPGAVSTLRDLVDVRFALVAGADPFQREWLMR 87  
 58 -----IVRVSVPTMDAAE-----AFKLIKQ-QVNVPLVADIHEDYRIA-- 94  
 88 VWEIDRIHELPPLTGLIDIALMDLAGRYGEGEPVRLGFEREIPAYASTSTFSSVAEF 147  
 95 LKVAEYGVDCIRINPGNIGNEERIRRMVVDACRDNIPRIGVNASLEKDLQEKY----G 150  
 148 LDVAD---QCLALRYGI-----KLHANGDARRDAELCLALRDHVG---PDVPLMTDSAG 197  
 151 EPTPOL-LESAMRHVDHLDRLNFDQPKVSKASDVFLAVESRYLAKQIDQPLHGTITE 209  
 198 FDLPLDAIRLGRALSEADYL-----WYEPIRE---FSISAYORLAEADVPLIVAEYS 247  
 210 AGG-----ARSAVSAIGLIGLISEGIGDITLRVSLADPVEIIVGPDILKSLIRS 262  
 248 DGAMNNADFTIRAGAAFGVAGCTTLRGITGAKRTALHADAER-----LRAEV 296  
 263 RGI---NFACPTGSRQEF--DVIGTVNALFORLEDITTPMDVSIIGCVNGPEALVST 317  
 297 HGSIDIPNHLCMAISNTTYESLTVSVAVVERHVD-----DQGLVHAPA--GPGIALPLD 350  
 318 LGVTGK 325  
 351 FG--GNE 356

RESULT 12  
 US-09-815-242-10969

Sequence 10969, Application US/09815242  
 Patent No. US20020061569A1

GENERAL INFORMATION:  
 APPLICANT: Haselbeck, Robert  
 APPLICANT: Ohlsen, Kari L.  
 APPLICANT: Zyskind, Judith W.  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John D.  
 APPLICANT: Carr, Grant J.  
 APPLICANT: Yamamoto, Robert T.  
 APPLICANT: Xu, H. Howard  
 TITLE OF INVENTION: Identification of Essential Genes in  
 TITLE OF INVENTION: Prokaryotes  
 FILE REFERENCE: ELITRA.011A  
 CURRENT APPLICATION NUMBER: US/09/815,242  
 CURRENT FILING DATE: 2001-03-21  
 PRIOR APPLICATION NUMBER: 60/191,078  
 PRIOR FILING DATE: 2000-03-21  
 PRIOR APPLICATION NUMBER: 60/206,848  
 PRIOR FILING DATE: 2000-05-22  
 PRIOR APPLICATION NUMBER: 60/207,727  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: 60/242,578  
 PRIOR FILING DATE: 2000-10-23  
 PRIOR APPLICATION NUMBER: 60/253,625  
 PRIOR FILING DATE: 2000-11-27  
 PRIOR APPLICATION NUMBER: 60/257,931  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: 60/269,308  
 PRIOR FILING DATE: 2001-02-16  
 NUMBER OF SEQ ID NOS: 14110  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 10969  
 LENGTH: 504  
 TYPE: PR  
 ORGANISM: Haemophilus influenzae  
 US-09-815-242-10969

Query Match 5.1%; Score 96; DB 9; Length 504;  
 Best Local Similarity 22.0%; Pred. No. 0.51;  
 Matches 85; Conservative 59; Mismatches 134; Indels 108; Gaps 21;

```

QY 36 TRTDEAVAFVNOJKALER-----VGADIVRSVPTMDAAE---AFKLIQGVWVPL 83
Db 24 TRTRD--ALVADPEGLTRRCKYGHANHAIVDEYIVDTGIDGTEEGVEEKMAEOSLIAD 81
QY 84 VADIH--DYRIALKVAEYV-DCLIRINPGNIGNEERIRMV-----DCARDK 128
Db 82 EADIVLEIVDARGGLUADIGIANTYR-----QONKITYVANKRDTGDIDASHCEFY 135
QY 129 NIP----RIGVNAAGLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKYSKASD 184
Db 136 QLGIGIEIQIASOGGVTQIMEQVLA-----FAKKENNDENDRSTSEEQDEMEQEPD 190
QY 185 VFLAVESYRLANQIDQPL-----HIGTENGARSAVSAIGLILSEIGD----- 234
Db 191 -FDSSEPTALIDDALEEBEODKNKIKAIYV-----RPNGKSTLTRLIGEDRVY 242
QY 235 -----TLRYSIADAVEEIKVFPDILSLIRSRGINFICPSCRSROEFPVIGTVALE 288
Db 243 VEDMPGTRNSIYI-PMERDGOQYLLIDVAGRKRGKHYLA-----EKSEVITTLQIQ 296
QY 289 QR-----LEDITPMDSIIGCVVNGPEALVSTLGVTGNKKSGLYEDGVKRDRLD 340
Db 297 DANVVLITIDARENISDODLSLGLFLNA-GRSLIYV-----NKMDELDD--VKDRV- 347
QY 341 NNDMIDQLEARIKASQOLDEARRID 366
Db 348 -----KSELD--RRLD 356

```

```

RESULT 13
US-10-260-877-70
: Sequence 70, Application US/10260877
: Publication No. US20030021813A1
: GENERAL INFORMATION:
: APPLICANT: Abbott Laboratories
: APPLICANT: Chovan, Linda E.
: APPLICANT: Hessler, Paul E.
: APPLICANT: Reich, Karl A.
: TITLE OF INVENTION: ESSENTIAL BACTERIA GENES AND GENOME
: TITLE OF INVENTION: SCANNING IN HAEMOPHILUS INFLUENZAE FOR THE IDENTIFICATION OF
: FILE REFERENCE: 6565, US, pl
: CURRENT APPLICATION NUMBER: US/10/260,877
: CURRENT FILING DATE: 2002-09-30
: PRIOR APPLICATION NUMBER: US/09/649,145
: PRIOR FILING DATE: 2000-08-25
: NUMBER OF SEQ ID NOS: 137
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 70
: LENGTH: 504
: TYPE: PRF
: ORGANISM: H. influenzae
: US-10-260-877-70

```

Query Match	5.1%;	Score 96;	DB 15;	Length 504;
Best Local Similarity	22.0%;	Pred. NO. 0.51;		
Matches	85;	Conservative	59;	Mismatches 134;
				Indels 108;
				Gaps 21.

```

0Y 36 TRTDDAVALVNDIKALER-----VGADIVKVSPTTMDAAE---AFKLKIQOVNPL 83
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24 TRTRD--ALVADEPGITRRKRYGKSHAHIAGYDFEYIDTGGIDGTEEGVEEKMAEGLSLAID 81
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
0Y 84 VADIHF---DYIALKVAERGV-DCLRINPGNIGNEERIRNV-----DCARDK 128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 82 EADIVLELVDAAGSLAADIGIANYL-----ORONKITVYVANKRTGDIDADSHCAEFY 135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
0Y 129 NPI-----RIGVNASLEKDLQEKGEPTPALLESAMRHVDLDRINPOQFVSKVAST 184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136 QLGISEIEGIIAASOGRGVYOLMEQVLAP-----FAEKKENADENDRTSEEDQEMEQEFD 190
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
0Y 185 VFLAVSRYLLAKOIDOPL-----HLGITAGCARSGAVKSAIGLILLSEGTGD---- 234
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

[illegible]

```

RESULT 14
US-09-815-242-5590
? Sequence 5590 Application US/09815242
? Patent No. US20020061569A1
? GENERAL INFORMATION:
?   APPLICANT: Haselbeck, Robert
?   APPLICANT: Ohlsen, Karl L.
?   APPLICANT: zyskind, Judith W.
?   APPLICANT: Wall, Daniel
?   APPLICANT: Trawick, John D.
?   APPLICANT: Carr, Grant J.
?   APPLICANT: Yamamoto, Robert T.
?   APPLICANT: Xu, H. Howard
?   TITLE OF INVENTION: Identification of Essential Genes In
?   TITLE OF INVENTION: Prokaryotes
?   FILE REFERENCE: ELITRA.011A
?   CURRENT APPLICATION NUMBER: US/09/815,242
?   CURRENT FILING DATE: 2001-03-21
?   PRIOR APPLICATION NUMBER: 60/191,078
?   PRIOR FILING DATE: 2000-03-21
?   PRIOR APPLICATION NUMBER: 60/206,848
?   PRIOR FILING DATE: 2000-05-23
?   PRIOR APPLICATION NUMBER: 60/207,727
?   PRIOR FILING DATE: 2000-05-26
?   PRIOR APPLICATION NUMBER: 60/242,578
?   PRIOR FILING DATE: 2000-10-23
?   PRIOR APPLICATION NUMBER: 60/253,625
?   PRIOR FILING DATE: 2000-11-27
?   PRIOR APPLICATION NUMBER: 60/257,931
?   PRIOR FILING DATE: 2000-12-22
?   PRIOR APPLICATION NUMBER: 60/269,308
?   PRIOR FILING DATE: 2001-02-16
?   NUMBER OF SEQ ID NOS: 14110
?   SOFTWARE: FastSeq for Windows Version 4.0
?   SEQ ID NO 5590
?   LENGTH: 290
?   TYPE: PRT
?   ORGANISM: Staphylococcus aureus
US-09-815-242-5590

```

Query Match	5.1%;	Score 94.5;	DB 9;	Length 290;
Best Local Similarity	21.68;	Pred. NO. 0.3;		
Matches 63;	Conservative 35;	Mismatches 106;	Indels 87;	Gaps 12

```

QY 27 PLVDSMTNRTTDEAT-----VNOIALERGVADIVRSVPMDAAEFKLIKQOVN 81
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 60 PKIVEYVMAVNSIPYMAKARIGHTEARVLEAMGVYIDSESVLTPEDEEHKRKQFTV 119
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 82 PLVADIHFPYRALKVAEYVDCLRINGNIGNEERIRAVVDCARKNIRIRGVNAGSL 141
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 120 PFVCGGR-----NIGEAR-----RIGEGAML 142
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 142 EKDLQEKYGEPPQALLSARHVDHDLRFDPKYSVAKSVFLVAVESYRLLAQIDQ 201
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 143 RTR-----GEPATGNIVEA-----VRHMQVNSEYSRLLVMNDEIMT-----FARDIGA 187
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 202 PLHIGTTEAGRSAAVNSATGIGLLSEIGD-----TLKVSIAADVEEIKTGFIDLS 257
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

Db 188 PYELIKOIKDNGRLPYVNFAG-----GVATPDAAIMELGADV---FVSGGIFKS 237

Job time : 58 secs

OY 258 LRIIRSGINFACPTGSRQEDFVIG-----TVN--ALEORLED 293  
Db 238 -EDPEKFAKAIYQATTHYQDVELIGRLASDLGTAMKGLIDINQLSLEERMGE 287

## RESULT 15

US-09-815-242-12428

Sequence 12428, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Karl L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

PRIOR FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-28

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-26

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 12428

LENGTH: 295

TYPE: PRT

ORGANISM: Staphylococcus aureus

US-09-815-242-12428

Query Match 5.18; Score 94.5; DB 9; Length 295;  
Best Local Similarity 21.63; Pred. No. 0.31;

Matches 63; Conservative 35; Mismatches 106; Indels 87; Gaps 12;

OY 27 PIAVOSMTNTRTDEAT-----VNOIKALEVSGADIVRSVPTMDAAEFKIKQVNV 81

Db 65 PKIVEVNNNAVSIPYAKARKIGHITAEVLEAMGVYIDSESVLTTPADEYHLRKDQFTV 124

OY 82 PLVADIHFDYRLAKVAEYGVCLINPQNGNEERIRMVVCCARDKNIPIRIGVAGSL 141

Db 125 PFVCGCR-----NLGEAR-----RIGEGAML 147

OY 142 EKDLQKRGERTPOARESSAMRHVDLRLNFDQFVSKASDVFLAVESYRLAKQIDQ 201

Db 148 RTK-----GEPGTGNIVEA---VHRMQVNSEVSRITVMNDELMT-----FAKDIGA 192

OY 202 PLHLGITBAGARSGAVKSAIGLILSEGID---TLRVSADPYEIKVGFILKS 257

Db 193 PYELIKOIKDNGRLPYVNFAG-----GVATPDAAIMELGADV---FVSGGIFKS 242

OY 258 LRIIRSGINFACPTGSRQEDFVIG-----TVN--ALEORLED 293

Db 243 -EDPEKFAKAIYQATTHYQDVELIGRLASDLGTAMKGLIDINQLSLEERMGE 292

Search completed: August 12, 2003, 10:07:05

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 14, 2003, 21:47:30 ; Search time 4626 Seconds  
(Without alignments)  
3289.749 Million cell updates/sec

Title: US-09-921-992-78  
Perfect score: 1866  
Sequence: 1 MHNQAPIQRKRRSTRIVYGVN.....RAKASQDEARIDVQGVK 372

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues  
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xih  
-O=/cgn2\_1/USPTO.spool/US09921992/runat.12082003.094911.17364/app\_query.fasta.1.519  
-DB=GenEmbl -OFMT=fasta -SUFFIX=rge -MIMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pc -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORTEXT -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09921992.ccgcn.1.1.3508.@runat.12082003.094911.17364 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_jmu.\*  
20: em\_jun.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*

29: em\_vl.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rtd.\*  
36: em\_htg\_man.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htg\_hum.\*  
40: em\_htg\_mus.\*  
41: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1866	100.0	1119	1 AY033515	AY033515 Escherich
2	1866	100.0	1119	6 AX036302	AX036302 Sequence
3	1866	100.0	1119	6 AX038912	AX038912 Sequence
4	1866	100.0	1119	6 AX050487	AX050487 Sequence
5	1866	100.0	1119	6 AX393943	AX393943 Sequence
6	1866	100.0	1697	1 ECGCPE	X64451 E.coli gcpe
7	1866	100.0	11521	1 AE005481	AE005481 Escherich
8	1866	100.0	13176	1 AE000338	AE000338 Escherich
9	1866	100.0	296827	1 AP002551	AP002551 Escherich
10	1866	100.0	300059	1 AE016764	AE016764 Escherich
11	1835	99.4	10225	1 AE015271	AE015271 Shigella
12	1835	99.4	290380	1 AE016987	AE016987 Shigella
13	1840	98.6	23647	1 AE008814	AE008814 Salmonella
14	1831	98.1	145050	1 AL627275	AL627275 Salmonella
15	1831	98.1	300431	1 AE016835	AE016835 Salmonella
16	1691	90.6	1010	6 AR027993	AR027993 Sequence
17	1691	90.6	1010	6 AR208328	AR208328 Sequence
18	1691	90.6	1010	6 BD084908	BD084908 Methods f
19	1645.5	88.2	301235	1 AP005075	AP005075 Vibrio pa
20	1631.5	87.4	301442	1 PS067933	PS067933 Providencia
21	1624.5	87.1	1345	6 AR027991	AR027991 Sequence
22	1624.5	87.1	1345	6 AR208326	AR208326 Methods f
23	1624.5	87.1	1345	6 BD084906	BD084906 Methods f
24	1624.5	87.1	1345	6 AE013738	AE013738 Yersinia
25	1622.5	87.0	204050	1 AJ414154	AJ414154 Yersinia
26	1622.5	86.9	11824	1 AE004161	AE004161 Vibrio ch
27	1621.5	85.9	10507	1 U32721	U32721 Haemophilus
28	1603.5	85.9	110000	6 AR274513_03	AR274513_03 Continuation (4 of
29	1593.5	85.7	11268	1 AE006237	AE006237 Pasteurel
30	1593.5	85.7	1079	6 AR027992	AR027992 Sequence
31	1593.5	85.4	1079	6 AR208327	AR208327 Sequence
32	1593.5	85.4	1079	6 BD084907	BD084907 Methods f
33	1593.5	84.9	1079	6 AE015769	AE015769 Sequence
34	1584.5	84.9	10607	1 AE015769	AE015769 Sequence
35	1372.5	73.6	17148	1 AE004798	AE004798 Shewanella
36	1353	72.5	886	6 AR027995	AR027995 Pseudomon
37	1353	72.5	886	6 AR208330	AR208330 Sequence
38	1353	72.5	886	6 BD084910	BD084910 Methods f
39	1341.5	71.9	300861	1 AE016777	AE016777 Pseudomon
40	1335.5	71.6	311249	1 AE016860	AE016860 Pseudomon
41	1285.5	68.9	9608	1 AE014104	AE014104 Buchnera
42	1282	68.7	110000	6 BD061520_3	BD061520_3 Continuation (4 of
43	1282	68.7	347550	1 AP001118	AP001118 Buchnera
44	1226.5	63.7	349174	1 AB063522	AB063522 Wiggleswo
45	1186	63.6	4775	1 D90881	D90881 E. coli geno

RESULT 1

## ALIGNMENTS

AY033515 1119 bp DNA linear BCT 26-DEC-2001  
 LOCUS Escherichia coli 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate  
 DEFINITION synthase (ispG) gene, complete cds.  
 ACCESSION AY033515  
 KEYWORDS AY033515.1 GI:17978527  
 SOURCE

ORGANISM Escherichia coli  
 Escherichia coli  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 1119)  
 Hecht, S., Eisenreich, W., Adam, P., Amslinger, S., Kis, K., Bacher, A.,  
 Arigoni, D., and Rohdich, F.  
 Studies on the nonmevalonate pathway to terpenes: The role of the  
 Gcpe (ispG) protein  
 Proc. Natl. Acad. Sci. U.S.A. 98 (26), 14837-14842 (2001)

JOURNAL 2 (bases 1 to 1119)  
 Rohdich, F., Hecht, S., Adam, P., Bacher, A. and Eisenreich, W.  
 Direct Submission  
 Submitted (27-APR-2001) Institut fuer Organische Chemie und  
 Biochemie, Lehrstuhl III, Technische Universitaet Muenchen,  
 Lichtenberg Strasse 4, Garching D-85747, Germany

FEATURES  
 Location/Qualifiers  
 1..1119  
 /organism="Escherichia coli"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:562"

gene 1..1119  
 /gene="ispG"  
 1..1119  
 /gene="ispG"  
 /function="converts 2C-methyl-D-erythritol  
 2,4-cyclo-diphosphate into 1-hydroxy-2-methyl-2-(E)-butenyl  
 4-diphosphate, last but one step of the  
 mevalonate-independent pathway for isoprenoid  
 biosynthesis"

/note="ispG"  
 /codon\_start=1  
 /transl\_table=11  
 /product="1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate  
 synthase"  
 /protein\_id="AAK53460.1"  
 /db\_xref="GI:17978528"

/translation="MHNOAPTORRKRSTRIYVNPFGDGPAPIAVOSMTNRTTDEAT  
 VNQIKALERTYGGADYIRVSVPTMDAARAKLIKQOVNVPVLVDIHEDYRIATKVAEYGV  
 DCLRTNPGNIEERIKRMVVDCAKDKNIPIRIGVAGSLKEDLDKEEPTPOALLES  
 AMRVYDHLDRINFDQFKVSVKASDVLAVESYRLAKIDPLHLGITEAGARGAV  
 KSAIGLIEGSEIGDITLRSVLAADPEVEIKVGEIDILSLRISRGINFACPTCSRQ  
 EFDVIGTVALLEORLEIDITPMDSVIGICVVGPEALVSTIGTSGNKKSGLYEDGV  
 RKDRDNDMDIDQLEARIKAKSOLDEARRIDVQVER"

BASE COUNT 272 a 280 G 315 g 252 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 9, 1e-136 Length: 1119  
 Score: 1866.00 Matches: 372  
 Percent Similarity: 100.003 Conservative: 0  
 Best Local Similarity: 100.003 Mismatches: 0  
 Query Match: 100.003 Indels: 0  
 DB: 1 Gaps: 0

US-09-921-992-78 (1-372) x AY033515 (1-1119)

QY 1 MethASaNGlAlaBPoILleglInaRgaYlYsSerThraRgILeYrValGlyASnaVl 20  
 Db 1 ATGCATTAACCGAGCTGCAATCAAGTAAATAATCAACGTAATTACCTTGGGAATG 60  
 QY 21 ProILleglYASpGLYAlaProILlealValGlnSerMetThraSnaThraRgThraSp 40  
 Db 61 CCGATTGGCGATGCTGCTCCATCGCGCGTACAGTCCATGACCAATACGCGTACGACAGAT 120

QY 41 ValGluAlaThrValASnGlnILeYsAlaLeuGluARValGlyAlaAspILeValArg 60  
 Db 121 GTGGAACAAACGGTCAATCAATCAAGCGCTGGAAACGGCTGATATCGTCCT 180  
 QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuILeYsGlnValaAn 80  
 Db 181 GTCTCCGTACCGAGATGAGCGCGGCAAGACCGCTCAAACTCATCAACAGCGGTTAAC 240  
 QY 81 ValProLeuValAlaAspILeHisPheAspTYrArgILeAlaLeuYsValaGluYr 100  
 Db 241 GTGCGGTGGTGGTGCATCACTCCACTTCGACTATCGCATTCGCTGAAGTAGGGAATAC 300  
 QY 101 GlyValAspCysLeuARgILeAsnProGlyASnILeGlyASnGluGluArgMet 120  
 Db 301 GGGCTGATGTGTGCGCTATTACCTGGCAATATCGTAAATAGACGCTATTCCCATG 360  
 QY 121 ValValAspCysAlaAlaGAspYsAsnILeProILeArgILeGlyValaAsnAlaGlySer 140  
 Db 361 GTGGTTGACTGTGGCGCGATAAACATTCGCCGATGCTTAAGCCGCGATCG 420  
 QY 141 LeuGluLysAspLeuGlnGluLysTYrGlyGluProThrProGlnAlaLeuGluSer 160  
 Db 421 CTGCAAAAAGATCTGCAGAAAAGTATGGGGAACCGACGCGAGGCTTGCGTAATCC 480  
 QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180  
 Db 481 GCCATGCGCGCATGTGTATCATCTCGATCGCTGAACCTTCATCATGTTCAAAAGTCAGCTG 540  
 QY 181 LysAlaSerAspValAlaPheLeuAlaValGluSerTYrArgLeuAlaLysGlnILeAsp 200  
 Db 541 AAAGCGTGTAGCGTCTTCTCGCTGTTGAGCTTATGTTTGGCGCAAAACAATGAT 600  
 QY 201 GlnProLeuHisLeuGlyILeThrGluAlaGlyAlaArgSerGlyAlaValLysSer 220  
 Db 601 CAGCGGTGATCGATCGGGATCACCAAGCGGCTGGCGGACGCGGAGCAAGTAAATCC 660  
 QY 221 AlaILeGlyLeuGlyLeuLeuLeuSerGlnGlyILeGlyAspThraLeuARgValSerLeu 240  
 Db 661 GCCATTGGTTTAAAGTCTGCTGCTCTGTGAAGGCAATCGGCGACGCTGCCGCTATCCTG 720  
 QY 241 AlaAlaAspProValGluGluILeLysValGlyPheAspILeLysSerLeuARgILe 260  
 Db 721 GCGGCGCATTCGCTCGAAGAGATCAAGTGGTTTCGATATTGTAATCGCTTCGTATC 780  
 QY 261 ArgSerArgILeAsnPheILeAlaCysProThrcysSerArgGlnGluPheAspVal 280  
 Db 781 CGTTCGCGAGGTATCAACCTTCATCGCTCGACCTGTCGCTCGCAAGAAATTTGATGTT 840  
 QY 281 ILeGlyThraValaAsnAlaLeuGlnARgLeuGluAspILeILeThrProMetAspVal 300  
 Db 841 ATCGGTACAGTTAACGCGCTGGAGCAACGCTCGAAGATATCATCATCCGATGAGAGTT 900  
 QY 301 SerILeILeGlyCysValaValaAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320  
 Db 901 TCGATTATCGGCTGGTGGTGAATGCGCGGCTGAGCGCGCTGTTTACACTCGCGGCTC 960  
 QY 321 ThrGlyGlyAsnLysLysSerGlyLeuTYrGlnLysPGLYValaArgLysAspArgLeuAsp 340  
 Db 961 ACCGCGCGCAACAAGAAAAAGCGCTCTATGAATGCGCGTGCACAAAGACCGCTGAC 1020  
 QY 341 AsnAsnAspMetILeAspGlnLeuGluAlaARgILeArgAlaLysAlaSerGlnLeuAsp 360  
 Db 1021 AACACGATATGATGACACACCTGAGCAGCAGCATGCTGTGGAAACGACAGTACCTGAGAC 1080  
 QY 361 GluAlaARgArgILeAspValGlnGlnValGluLys 372  
 Db 1081 GAAGCGCTCGAATTGACGTTACAGCAGGTTGAAAAA 1116

RESULT 2  
 LOCUS AX036302 1119 bp DNA linear PAT 16-NOV-2000  
 DEFINITION Sequence 29 from Patent EP1043403.  
 ACCESSION AX036302

```

VERSION      AX036302.1  GI:11225912
KEYWORDS
SOURCE       Escherichia coli
ORGANISM     Escherichia coli
REFERENCE    1
AUTHORS      Enterobacteriaceae; Escherichia.
TITLE        Novel method for identifying antibacterial compounds
JOURNAL      Patent: EP 1043403-A 29.11-OCT-2000;
              GPC AG GENOME PHARMACEUTICALS (DE)
FEATURES     Location/Qualifiers
              1. 1119
              /organism="Escherichia coli"
              /mol_type="genomic DNA"
              /db_xref="taxon:562"
BASE COUNT   272 a 279 c 317 g 251 t
ORIGIN
Alignment Scores:
Pred. No.:    9.1e-136      Length:      1119
Score:        1866.00      Matches:    372
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match:  100.00%      Indels:    0
DB:           6           Gaps:      0
US-09-921-992-78 (1-372) x AX036302 (1-1119)
QY      1 MethisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
DB      1 ATGCATACCAGGCTCCAAATTCACGTAGAAAATCAACACGTATTTACCTGGGAATTCG 60
QY      21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
DB      61 CCGATTGGCGATGGTCTCCATCCGCGGTACAGTCCATACCAATCCGCTACGACAGAC 120
QY      41 ValGlnAlaThrValAsnGlnIleLysAlaLeuGlnArgValGlyAlaAspIleValArg 60
DB      121 GTGGAGGACACGGTCAATCAATCAAGCGCTGGAACGGCTGGCTGATATCGCCCT 180
QY      61 ValSerValProThrMetAspAlaAlaGlnAlaPheLysLeuIleLysGlnGlnValAsn 80
DB      181 GATCGCTGACGACGATGACCGCGAGAGAGCGCTTCAACTATCAACAGCAGGTTAAC 240
QY      81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGlnTyr 100
DB      241 GTGCCGCTGGTGGTGACATCCACTTCGACTATCGCATTCGCGCTGAAGTACCGGATTC 300
QY      101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGlnGluArgIleArgMet 120
DB      301 GCGGTGATGTTGCTGCTGATTTAACTGCAATATCGGTAAAGAGCGTATTCGATG 360
QY      121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAlaAsnIleArgSer 140
DB      361 GTGGTTGACTGTGCGCGGATTAACCAATCCGATCCGATGGCGTTAACCCCGGATCG 420
QY      141 LeuGlnLysAspLeuGlnGlnLysTyrGlnLysIleProThrProGlnAlaLeuLeuGlnSer 160
DB      421 GTGGAAAAAGATCTCAAGAAAAAGTATGGCGAAGCCGACGCGCTGGTGGATTC 480
QY      161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
DB      481 GCCATGCGGTGATGATCATCTCGATCGCTGAACCTTGATCAAGTCAAGCGCTG 540
QY      181 LysAlaSerAspValPheLeuAlaValGlnSerTyrArgLeuLeuAlaLysGlnIleAsp 200
DB      541 AAAGCGTGAAGCTTCTCTCGCTGTGAGCTTATCGTTTGCTGGCAAAACAGATCAT 600
QY      201 GlnProLeuHisLeuGlnIleTyrHisGlnAlaGlyAlaArgSerGlyAlaValLysSer 220
DB      601 CAGCGGTGATCTGGGGATCACCGAAGCCGCTGGTGGCGCAGCGGGCGAGTAAATCC 660

```

```

QY      221 AlaIleGlyLeuGlyLeuLeuSerGlyGlyIleGlyAspThrLeuArgValSerLeu 240
DB      661 GCCATTGGTTTAGCTCTGCTGCTGTGTAAGGATCGGACACGCTGCGGATCGCTG 720
QY      241 AlaAlaAspProValGlnGlnIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
DB      721 GCGCGCATCCCGCTCCAAAGATCAAGATCAAGTGGTTTGATTTGAAATCGCTGCTATTC 780
QY      261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnLysPheAspVal 280
DB      781 CGTTCCGAGGAGATCAACTTCACTCCGCTCCCGACCTGTCCGCTAGGAATTTGATGTT 840
QY      281 IleGlyThrValAsnAlaLeuGlnGlnArgLeuGlnAspIleIleThrProMetAspVal 300
DB      841 ATCGGATCGGTTTACCGCTGAGCAACCGCTGGAGATATCATCATCCGATGACAGTT 900
QY      301 SerIleIleGlyCysValAlaAsnGlyProGlyGlnAlaLeuValSerThrLeuGlnVal 320
DB      901 TCGATTATCGCGTGGCTGGTGTGATATGCGCCAGGTGAGCGCTGTTCTTCACTCGCGCTG 960
QY      321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGlnAspGlyValArgLysAspArgLeuAsp 340
DB      961 ACCGGCGGACACAGAAAGACCGCTCTATGAAGATGGCGTCCGCAAAAGACGCTCGGAC 1020
QY      341 AsnAsnAspMetIleAspGlnLeuGlnAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
DB      1021 AACACAGATATATGACACCACTGGAGACGCACTTCGTGGAAAGCCAGTCACTGGTAC 1080
QY      361 GlnAlaArgArgIleAspValGlnGlnValGlnLys 372
DB      1081 GAAGCGCGCTGCAATTGACGTTACAGCTTACAGGTTGAAAAA 1116
RESULT 3
AX038912      1119 bp      DNA      linear      PAT 16-NOV-2000
LOCUS        AX038912      Sequence 29 from Patent W00061793.
DEFINITION   AX038912
ACCESSION    AX038912
VERSION      AX038912.1  GI:11228221
KEYWORDS
SOURCE
ORGANISM     Escherichia coli
              Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
              Enterobacteriaceae; Escherichia.
REFERENCE    1
AUTHORS      Loferer,H. and Jacobl,A.
TITLE        Novel method for identifying antibacterial compounds
JOURNAL      Patent: WO 0061793-A 29.19-OCT-2000;
              GPC BIOTECH AG (DE) ; LOFERER HANNES (DE) ; JACOBI ALEXANDER (DE)
FEATURES     Location/Qualifiers
              1. 1119
              /organism="Escherichia coli"
              /mol_type="genomic DNA"
              /db_xref="taxon:562"
BASE COUNT   272 a 279 c 317 g 251 t
ORIGIN
Alignment Scores:
Pred. No.:    9.1e-136      Length:      1119
Score:        1866.00      Matches:    372
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match:  100.00%      Indels:    0
DB:           6           Gaps:      0
US-09-921-992-78 (1-372) x AX038912 (1-1119)
QY      1 MethisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
DB      1 ATGCATACCAGGCTCCAAATTCACGTAGAAAATCAACACGTATTTACCTGGGAATTCG 60
QY      21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
DB      61 CCGATTGGCGATGGTCTCCATCCGCGGTACAGTCCATACCAATCCGCTACGACAGAC 120

```

QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60  
 DB 121 GTGACAGCAACGGGCAATCAATCAAGCGCGTGGAGCGCGTGGCGTGAATCCGCT 180  
 QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80  
 DB 181 GTATCCGTACCGAGGATGAGCGGCGAGACGGTTCAACTCATCAACAGCAGGTTAAC 240  
 QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100  
 DB 241 GTGCGCGGTGGGTGACATCCATCTGACATCGCATCGCATCGCGTGAAGAGTACGGAATAC 300  
 QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120  
 DB 301 GCGCTGCATGCTGCGGATTAACCTGGCAATATCGGTAATGAAGACCGTATTCGCATG 360  
 QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAlaAsnAlaGlySer 140  
 DB 361 GTGGTGAAGTGGCGCGGATTAACCTTCGATCCGATTCGCTTAACGCGGATCG 420  
 QY 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160  
 DB 421 CTGGAAGAAAGATCTGCAAGAAAGATGCGAGACCGCGCGCGGCTTCTGGAATCT 480  
 QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnProPheAspGlnPheLysValSerVal 180  
 DB 481 GCCATGCTCATGTTGATCATCTGCATCGCTGAACCTTCGATCGATTAAGTCAAGCGCTG 540  
 QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200  
 DB 541 AAGCGCTGAGCTTCTCTCGCTGAGTCTTATCGTTGCGGCAAAACAGATCGAT 600  
 QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220  
 DB 601 CAGCGCTGCAATCGGCGATCACCGCAACCGGTGTCGCGCGAGGGGCGAGTAAATCC 660  
 QY 221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240  
 DB 661 GCCATTGCTTGAAGTCTGCTGCTGCTGATGAGGCAACGCGCGATCGCTGATCGCTG 720  
 QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260  
 DB 721 GCGGCGGATCCGCTGGAAGATCAAACTCGGTTTCGATATTTGAAATCGCTGCGATTC 780  
 QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGluPheAspVal 280  
 DB 781 CATTGCGAGGAGATCACTTCATCGCTGCGCGACCTGTTCCGCTCAGAGAAATTTGATGT 840  
 QY 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300  
 DB 841 ATCGGTACGGTTAACGCGCTGAGCAACGCTGGAAGATATCATCATCCGATGAGCGTT 900  
 QY 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320  
 DB 901 TCGATTAATCGCTGCGGTGATGATGAGCCAGGTGAGCGCGCTTTCACACTCGGCGTC 960  
 QY 321 ThrGlyGlyAsnLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340  
 DB 961 ACCGCGCGCAACAGAAAGCGGCTCTATGAAGATGGCGTCCGCAAGACCGTCTGCAG 1020  
 QY 341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360  
 DB 1021 AACAAAGATATGATGACACGACGTGGAAGACGCAATCGTGCAGAAAGCCAGTACGTGCAG 1080  
 QY 361 GluAlaArgArgIleAspValGlnGlnValGluLys 372  
 DB 1081 GAAGCGCTGCGATTAAGCTTCAGCAGGTTGAAAAA 1116  
 RESULT 4  
 AX050487 1119 bp DNA linear PAT 12-JAN-2001  
 LOCUS AX050487  
 DEFINITION Sequence 1 from Patent WO0072022.

ACCESSION AX050487  
 VERSION AX050487.1 GI:12226691  
 KEYWORDS  
 SOURCE Escherichia coli  
 ORGANISM Escherichia coli  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.  
 REFERENCE  
 1 Jomaa, H.  
 Use of genes of the deoxy-d-xylose phosphate biosynthetic pathway for altering the concentration of isoprenoid  
 Patent: WO 0072022-A 1 30-NOV-2000;  
 Jomaa, Hassan (DE)  
 FEATURES  
 source location/Qualifiers  
 1..1119  
 /organism="Escherichia coli"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:562"  
 1..1119  
 /note="unnamed protein product"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAC21801.1"  
 /db\_xref="GI:12226692"  
 /translation="MHNQAPTRKSTRIVGVNPICDAPFIVGSKMTNRTDVEATVNCIKALERYGADIVRSVPTMDAAEFKIKQOVNPLVADHFDYRIALKYAETGVDCIRINGNIGNEERIIMVDCARDKNIPIRIGVNAISLEKLOERYGEPDQALGVAMRHVDHLRLNDQFVSVKASDVFVLYSLAKIDQPLHGLITFAGRSAGVKSATIGLILSEIGDRLRYSIADPVEEIKVGEFDILKSLRISRGITFACPTCSROEPDYITGVNALEQRLEDITTPMVSIIIGCVNPGRAIVSTLAVTGNNKKSGLYEDGV RKDRLDNDMDIDLEAIRAKASQLEDAKIDVQYER"  
 BASE COUNT 272 a 279 c 317 g 251 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 9 1e-136 length: 1119  
 Score: 1866.00 matches: 372  
 Percent Similarity: 100.00% conservative: 0  
 Best Local Similarity: 100.00% mismatches: 0  
 Query Match: 100.00% indels: 0  
 DB: 6 gaps: 0  
 US-09-921-992-78 (1-372) x AX050487 (1-1119)  
 QY 1 MethisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20  
 DB 1 ATGCATTAACCGAGCTCCATTCATCAAGTGAAGAAATCAACAGCTATTACGTTGGGAATGTG 60  
 QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40  
 DB 61 CGGATGGCGATGGTGTCCCTCCATCGCGCTACAGTCATGACCATACGCGTACGACAGAC 120  
 QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60  
 DB 121 GTGACAGCAACGGGCAATCAATCAAGCGCGTGGAGCGCGTGGCGTGAATCCGCT 180  
 QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80  
 DB 181 GTATCCGTACCGAGGATGAGCGGCGAGACGGTTCAACTCATCAACAGCAGGTTAAC 240  
 QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100  
 DB 241 GTGCGCGGTGGGTGACATCCATCTGACATCGCATCGCATCGCGTGAAGAGTACGGAATAC 300  
 QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120  
 DB 301 GCGCTGCATGCTGCGGATTAACCTGGCAATATCGGTAATGAAGACCGTATTCGCATG 360  
 QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAlaAsnAlaGlySer 140  
 DB 361 GTGGTGAAGTGGCGCGGATTAACCTTCGATCCGATTCGCTTAACGCGGATCG 420  
 QY 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160



```

|||||
Db 421 CTGGAAGAAATCGCAGAAAGTATGGCAACCGACGCGCGGCTTGCGATCT 480
QY 161 Almetarqhisvalaspshisluaspargleuasnpheaspplnphelyvalserval 180
Db 481 GCCATGGGTATGTTGATCTCATCTCGATCGCTCAACTCGATCGATTCAAGTCAGCGTG 540
QY 181 Lysalaseraspvalpheleualavalglusertyrargleuualalysglinileasp 200
Db 541 AAAGCGTGCAGCGTCTCCCTCGCTTGATCTTATCGTTTGCTGCGAAAACGATCGAT 600
QY 201 Gluproteinhlisleuglylierthrglualaglylalaargserglyalavalysser 220
Db 601 CACCGCGTGCATCTGGGATCCAGCAACCGGTGTGCGCGGACGCGGCGAGTAAATCC 660
QY 221 AlaileglyleuglyleuleuSerGluGlylierghlyaspshreleuargvalserleu 240
Db 661 GCCATTGGTTAGTGTCTGCTGCTGCTGGAAGGATCGGCGACACGCTGCGCGTATCGCTG 720
QY 241 AlaAlaasprovalglugluilelysvalglypheaaplleuleuysSerleuArgile 260
Db 721 GCGCGCGATCCGCTCGAAGATCAAGTGGTTGATTTGAAATCGCTGCGATC 780
QY 261 ArgSerargglylleasnphelealeagysprothrCyserargglngluipheaspval 280
Db 781 CGTTCGGAGGATCACTTCATCGCTGCGCGCGATCGCTGCGTCAAGGATTTGATGTT 840
QY 281 IleGlyThrValAsnAlaLeuGluGluArgLeuGluAspIleIleThrProMetAspVal 300
Db 841 ATCGGATACGGTTAAACGCGGTGAGCAACGCGTGAAGATATCATCATCGATGACGCTT 900
QY 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrleuGlyVal 320
Db 901 TCATATTATCGCGCTGCTGGAATGGCCCGAGTGAGCGCGCTGCTTCACACGCGCGCTC 960
QY 321 ThrGlyValAsnLysSerGlyLeuTyrgluaspGlyValArgLysAspArgleuAsp 340
Db 961 ACCGCGCGCAACAGAAAGCGCGCTCTATGAGATGGCGTGGCAAGACGCTGGAC 1020
QY 341 AsnAsnAspMetIleaspGluLeuGluAlaArgIleArgAlaLysAlaSerGluLeuAsp 360
Db 1021 AACCAAGATATGATCATCACACGCTGGAAGCAGCATTCGTGCGAAACCGATCGCTGAC 1080
QY 361 GluAlaArgArgIleaspValGluGluValGlyLys 372
Db 1081 GAAGCGCGTGCAGATTGACGTTCAAGGTTGAAAAA 1116

```

```

/protein_id="CAD29065.1"
/db_xref="GI:19701905"
/translation="MHQAPIDRRSTRIVGNVPIGDAPIAVSMNTRTTDEAT
VNOIKALEVAGADIVAVSVPTDAAEAFILIQOVNVPVADIHEDYRATLVAEYV
DCIRIPGNIGNEERIIMVDCARDKNIPRIIGVAGSLEKDOEYGEPTQALLES
ARHVDHLDRLNEDOFKSVKASDVLAESYRLAKOIDOLHIGITPAGARGSAV
KSAIGLILSISGIDPTLRLAADPVEKTRPGFKSLIRSRGINIACPTCSRO
FEDVIGTVALASORLEDITTPMDVSLICEGVVNGPEALYSTLGVTGNNKSLGLEDGV
KRDLDNMIDTQLEAIRAKASQDEARRIDVQVER"
BASE COUNT 272 a 279 c 317 g 251 t
ORIGIN
Alignment Scores:
Pred. No.: 9,1e-136 Length: 1119
Score: 1866.00 Matches: 372
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
US-09-921-992-78 (1-372) x AX393943 (1-1119)
QY 1 MethAsnGlnAlaProIleGluArgArgLysSerThrArgIleTyrgluAlaGlyAsnVal 20
Db 1 ATGCATTAACCAAGGCTCCCAATTCACGTAGAAATCAACACGATATTAGCTGGGAATG 60
QY 21 ProIleGlyAspGlyValaProIleAlaValGluInsMetThrAsnThrArgThrAsp 40
Db 61 CCGATTGGCGATGGGTGCTCCATCGCGCTACAGTCCATGCAAAATACCGGTACAGAC 120
QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
Db 121 GTGCAAGCAACGGTCAATCAAAATCAAGGCGTGAAGCGGTGGCGCTGATTCGCTCCT 180
QY 61 ValSerValProThrMetAspAlaAlaGlyAlaPheLysLeuIleLysGlnValAsn 80
Db 181 GATATCCGTACCAAGATGAGCGCGGAGAGGTTCAAACTCATCAACACGAGGTTAAC 240
QY 81 ValProLeuValAlaAspIleHisPheAspTyrglyIleAlaLeuLysValAlaGluTyrg 100
Db 241 GTGCGCGTGTGGTGCACATCCACTTCGACTTCGCTGCGCTGAAAGTACGGAAATG 300
QY 301 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
Db 361 GTGCGTGAATGTCTGCGATTAACCTGCAATATCGGAATAGAAAGAGGTATGCGANTG 360
QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaLysSer 140
Db 361 GTGTTGACTGTGCGCGCATTAACATTCGATCCGATTCGATTCGCGTTAACCGCGATCG 420
QY 141 LeuGluLysAspLeuGluGluTyrglyGluProThrProGluAlaLeuLeuGluSer 160
Db 421 CTGGAAGAAATCTCCAGAAAGATATGCGCAACGACGCGCGCTGCTGGAATCT 480
QY 481 Almetarqhisvalaspshisluaspargleuasnpheaspplnphelyvalserval 180
Db 481 GCCATGGGTATGTTGATCTCATCTCGATCGCTCAACTCGATCGATTCAAGTCAGCGTG 540
QY 541 Lysalaseraspvalpheleualavalglusertyrargleuualalysglinileasp 200
Db 541 AAAGCGTGCAGCGTCTCCCTCGCTTGATCTTATCGTTTGCTGCGAAAACGATCGAT 600
QY 201 Gluproteinhlisleuglylierthrglualaglylalaargserglyalavalysser 220
Db 601 CACCGCGTGCATCTGGGATCCAGCAACCGGTGTGCGCGGACGCGGCGAGTAAATCC 660
QY 221 AlaileglyleuglyleuleuSerGluGlylierghlyaspshreleuargvalserleu 240
Db 661 GCCATTGGTTAGTGTCTGCTGCTGCTGGAAGGATCGGCGACACGCTGCGCGTATCGCTG 720
QY 241 AlaAlaasprovalglugluilelysvalglypheaaplleuleuysSerleuArgile 260
Db 721 GCGCGCGATCCGCTCGAAGATCAAGTGGTTGATTTGAAATCGCTGCGATC 780

```

CDS

```

/note="unamed protein product"
/codon_start=1
/transl_table=11

```

```

FEATURES
    source
        1..1119
            /organism="Escherichia coli"
            /mol_type="genomic DNA"
            /db_xref="taxon:562"
            1..1119

```

REFERENCE

AUTHORS

TITLE

JOURNAL

Monsanto

Patent: WO 0212478-A 3 14-FEB-2002;

Technology LLC (US)

Location/Qualifiers

1..1119

/organism="Escherichia coli"

/mol\_type="genomic DNA"

/db\_xref="taxon:562"

1..1119

/note="unamed protein product"

/codon\_start=1

/transl\_table=11

OY 261 ArgSerArgGlyIleAsnPhelIalacysProthrcysSerArgGlnGlnPheAspVal 280  
 Db 781 CGTTGGGAGGAGCAACTTCATCGCTGCCGCCGCTGCTCGCTAGGAATTTGATGTT 840  
 OY 281 IlleglyThrValAsnAlaLeuGlnGlnArgLeuGlnAspIleIleThrProMetAspVal 300  
 Db 841 ATCGGTACGGTTAACGGCTGAGACACGCTGGAAGATATCATCCTCCGATGAGCGTT 900  
 OY 301 SerIleIlleglyCysValValAsnGlyProGlyGlnAlaLeuValSerThrLeuGlyVal 320  
 Db 901 TCGATTATCGGCTCGTGTGAATGCCAGGTAGCGCGCTGTTCTACACTCGCGCTC 960  
 OY 321 ThrGlyGlyAsnLysLysSerGlyLeuThrGlnAspGlyValArgGlyAspArgLeuAsp 340  
 Db 961 ACCGGGGGCAACAAAGAAAGCGCGCTCTATGAAATGGCGGTGCGAAGACCGTCTGAC 1020  
 OY 341 AsnAsnAspMetIleAspGlnLeuGlnAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360  
 Db 1021 AACACGATATGATCGACACGCTGGAAGACGACGATTCGTGCGAAGCCAGTCAGCTGAC 1080  
 OY 361 GluAlaArgArgIleAspValGlnGlnValGlnLys 372  
 Db 1081 GAAGCGCGTCGAATTGACGTTCAAGCGTTGAAAAA 1116  
 RESULT 6  
 ECGCPE 1697 bp DNA linear BCT 08-DEC-1993  
 LOCUS E.coli gcpe gene.  
 DEFINITION X64451.1 S43432  
 ACCESSION X64451.1 GI:41540  
 VERSION gcpe gene.  
 KEYWORDS Escherichia coli  
 SOURCE Escherichia coli  
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Escherichia.  
 REFERENCE 1 (bases 1 to 1697)  
 AUTHORS Parker, J.  
 TITLE Direct Submission  
 JOURNAL Microbiology, S.I.U., Carbondale IL 62901, USA  
 REFERENCE 2 (bases 1 to 1697)  
 AUTHORS Baker, J., Franklin, D.B. and Parker, J.  
 TITLE Sequence and characterization of the gcpe gene of Escherichia coli  
 JOURNAL FEMS Microbiol. Lett. 94, 175-180 (1992)  
 COMMENT See also J01629 & M11843.  
 FEATURES  
 location/Qualifiers  
 1..1697  
 /organism="Escherichia coli"  
 /mol\_type="genomic DNA"  
 /strain="K12"  
 /db\_xref="taxon:562"  
 /sub\_clone="pSE411"  
 /cell\_line="JF298"  
 /clone\_lib="lambda transducing phage dguab-2"  
 1..508  
 /note="ORF"  
 /codon\_start=2  
 /transl\_table=11  
 /protein\_id="CAA45782.1"  
 /db\_xref="GI:41541"  
 /db\_xref="SWISS-PROT:P27434"  
 /translation="TTTDPATTSTPPASVDVTTANTQTTPAVTAPAPAVDPOONAVSP  
 SOANDTATPAPATATPAGAAPLPDAGVTPVADPNALVMTFTADCMLEVDAT  
 GKLESGMRKDGKNTLTGQAPYKFKIGAPAAVQIYOGKPDVLSRFTRTNQVALTL  
 NAEOSPAP"  
 535..1653  
 /gene="gcpe"  
 535..1653  
 /gene="gcpe"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA45783.1"

/db\_xref="GI:41542"  
 /translation="MENQAPIORRKRSTVYVNPIDGAPILAVOSMTNTRTDEAT  
 VNIKALERGALIVRSVPTMAAEFKILKQVNVPIYADHPRRIATKAEYGV  
 DCIRINPENGINEERIRNVDCARNDKPIRISVNNAGSLEKDLQETGEPYALLES  
 AMRHVHLDRNLNDOFRVSKADVFLAVSYLKLQIDQPHLITTEGGRSAV  
 KSAIGLILSEIGDILRVSLADPVEERIKVDFILKRSIRSRINIFACTPESRO  
 EFVDVIGVNALREQLEDITPMDSYIIGCVNPGPGLAVSTLGVTGNNKSGLYEDGV  
 RKRDLDNDMDIDLEARIIRAKASOLDEARIDVQVEK"  
 BASE COUNT 417 a 446 c 469 g 365 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1,586-135 Length: 1697  
 Score: 1866.00 Matches: 372  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 1 Gaps: 0  
 US-09-921-992-78 (1-372) x ECGCPE (1-1697)  
 OY 1 MethisAsnGlnAlaProIleGlnArgLysSerThrArgIleThrValGlyAsnVal 20  
 Db 535 ATGATTAACCAAGCGCTCCAAATTCAGTAGAAATCAACAGATTTACGTTGGGAATGTG 594  
 OY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40  
 Db 595 CCGATTGGCGATGTGTGCTCCATCGCGCTACAGTCAAGCAATACGCTAGACACAGAC 654  
 OY 41 ValGlnAlaThrValAsnGlnIleLysAlaLeuGlnArgValGlyAlaAspIleValArg 60  
 Db 655 GTCAAGCAACAGCGCAATCAATCAAGCGCTGGAACGCGTGGCGTGAATGTGCGT 714  
 OY 61 ValSerValProThrMetAspAlaAlaGlnAlaPheLysLeuIleLysGlnGlnValAsn 80  
 Db 715 GTATCCGTACCGAGAGATGAGACGCGCGCAAGCGTTCAACACATCAACAGACGTTAAC 774  
 OY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGlnTyr 100  
 Db 775 GTGCCGCTGGTGGCTGCATCCACTTCGACATTCGATTCGCGTGAAGAAAGTAGCGGAATAC 834  
 OY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGlnGlnArgIleArgMet 120  
 Db 835 GCGCGCATTTCTGCGATTAACCTGCGCAATTCGGTAAATGAAGACGATTCGCATG 894  
 OY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAlaAsnAlaGlySer 140  
 Db 895 GTGTTGACTGTGGCGCGGATAAAAACATTCGATTCGATTCGCGTAAACGCCGATACG 954  
 OY 141 LeuGlnLysAspLeuGlnGlnLysTyrGlyLeuProThrProGlnAlaLeuLeuGlnLys 160  
 Db 955 CTGGAAAAAGATCTGCAAGAAAGATATGCGCAACCGCGCGCGCGCTGCTGAATCT 1014  
 OY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180  
 Db 1015 GCCATGCTCATGTGATGATCTGATGCTGCAACTTCGATTCGATTCGATTCGATTCG 1074  
 OY 181 LysAlaSerAspValPheLeuAlaValGlnSerTyrArgLeuLeuAlaLysGlnIleAsp 200  
 Db 1075 AAAGCGTGTGAGCTTCCTCGCTGCTGATGATCTGCTGCTGCTGCTGCTGCTGCTG 1134  
 OY 201 GlnProLeuHisLeuGlyIleThrGlnAlaLeuGlyAlaArgSerGlyAlaValLysSer 220  
 Db 1135 CAGCGCTTGACATCTGGGGATACCGCAACCGCGTGGTGGCGCGCGCGCGGTAAATCC 1194  
 OY 221 AlaIleGlyLeuGlyLeuLeuLeuSerGlnGlyIleGlyAspThrLeuArgValSerLeu 240  
 Db 1195 GCCATTGCTTTAGCTCTCTGCTGCTGAAGCATGCGCGACGCTGCGGTATCGCTG 1254  
 OY 241 AlaAlaAspProValGlnGlnIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260  
 Db 1255 GCGGCGCATCGCGTGAAGAGATCAAGTCGTTTCGATTTTGAATTCGCTCGTATC 1314

QY 261 ArgSerArgGlyIleAsnPhelIleAlaCysProThrCysSerArgGlnIupheAspVal 280  
 Db 1315 CGTTCGGAGGATCACTTCATCCCTGCCGACCTGCGCTCAGGAATTTGANGTT 1374  
 QY 281 ILeGlyThrValAsnAlaLeuGlnIupheAspValIleLeuThrProMetAspVal 300  
 Db 1375 ATCGGTCGGGTAAACGGCTGGAGCAACGCTGGAAATATCATCATCTCGATGGAGCTT 1434  
 QY 301 SerIleIleGlyCysValValAsnGlyProGlyGlyAlaLeuValSerThrIleGlyVal 320  
 Db 1435 TCGATTATCGGCTCGCGGTGATGGCCAGGTGAGGCGCTGCTTCTACACTCGCGCTC 1494  
 QY 321 ThrGlyGlyAsnIleGlySerGlyLeuTyrGlyAspGlyValArgIleAspArgLeuAsp 340  
 Db 1495 ACCGGCGGCAACAGAAACGGCGCTTATGAAGATGGCGTGGCAACACCGCTGGAC 1554  
 QY 341 AsnAsnAspMetIleAspGlnIleuGlnAlaArgIleArgAlaIleAspGlnLeuAsp 360  
 Db 1555 AACACGATATGATCGACGAGCTGGAAACGCAATTCGTCGAAAGCAGTCAAGTGGAC 1614  
 QY 361 GluAlaArgArgIleAspValGlnGlnValGluLys 372  
 Db 1615 GAAGCGGCTGGAATGACGTTGACAGGTTGAAAA 1650

RESULT 7  
 AE005481/c 11521 bp DNA linear BCT 21-MAR-2001  
 LOCUS Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 100  
 DEFINITION  
 ACCESSION AE005481 AE005174  
 VERSION AE005481.1 GI:12516908  
 KEYWORDS  
 SOURCE Escherichia coli O157:H7 EDL933  
 ORGANISM Escherichia coli O157:H7 EDL933  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.

REFERENCE  
 AUTHORS Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Postel, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grobleck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A. and Blattner, F.R.  
 TITLE Genome sequence of enterohemorrhagic Escherichia coli O157:H7  
 JOURNAL Nature 409 (6819), 529-533 (2001)  
 MEDLINE 21074935  
 PUBMED 11206551  
 REFERENCE 2 (bases 1 to 11521)  
 AUTHORS Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Postel, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grobleck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A. and Blattner, F.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA  
 FEATURES  
 source location/Qualifiers  
 1..11521  
 /organism="Escherichia coli O157:H7 EDL933"  
 /mol\_type="genomic DNA"  
 /strain="EDL933"  
 /serotype="O157:H7"  
 /db\_xref="taxon:155864"  
 /note="enterohemorrhagic"  
 complement(86..1204)  
 /gene="gcpE"  
 /note="synonym: 23778"  
 complement(86..1204)  
 /gene="gcpE"  
 /function="orf; Unknown function"  
 /note="Residues 1 to 372 of 372 are 100.00 pct identical

to residues 1 to 372 of 372 from Escherichia coli K-12  
 Strain MG1655: B2515"  
 /codon\_start=1  
 /transl\_table=11  
 /product="orf, hypothetical protein"  
 /protein\_id="AG57625.1"  
 /db\_xref="GI:12516909"  
 /translation="MHNQAPIDRRKSTRIVGVNPIGDGAPITAVOSMTNRTDVEAT  
 VNQIKALERVADIVRVSVPDAAEAFILIQVNVPAVIDHPYRATLVAEIV  
 AMHVDLRLNFDQKSVKASDVLAESYRLKQIDQIDPLHGTITAGARSAV  
 KSAIVIGLLISGIDGLVSLAADPEVEIKGFDLKLRIKRSRINFIAQPTGSRQ  
 EEDVIGTVALERQLEDITIPMDVSIICGVNPGCALVSTLGVTGNKSSGLYEDGV  
 RKDRIDNNMIDLEARIKAKASQDEARIDVQVEK"  
 complement(1231..2244)  
 /gene="yfgA"  
 /note="synonym: 23779"  
 complement(1231..2244)  
 /gene="yfgA"  
 /function="putative membrane; Not classified"  
 /note="Residues 1 to 337 of 337 are 99.10 pct identical to  
 residues 1 to 337 of 337 from Escherichia coli K-12 Strain  
 MG1655: B2516"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative membrane protein"  
 /protein\_id="AG57626.1"  
 /db\_xref="GI:12516910"  
 /translation="NNTKATHQNEALTTGARLRNAREQLGSQAVARLEKSTV  
 RIEDKAPADLASTFRLRGYIRSVARLVIPBEELIPGLEKAPRAAKVAMQSESL  
 GRRKRKRDQMLMTFTWLVLVIVIGSGAMMDHRAQOEITTMADQSSXELSSQ  
 GQSVPLNTXTTDDPATSTSPASVDITATNTQTPVTAPAPAVDQOANVSPSANY  
 DTAATPAPATTPDGAAPLPIDQAGVTTPVADPAIVNNTADQMLVETDQTKLPS  
 SGWQRKDGMLNTGQAPYKIKIGAPRAAVQIQGQRPVLSRIFRINQVARTLNAQDS  
 PAQ"  
 complement(2529..3683)  
 /gene="yfgB"  
 /note="synonym: 23780"  
 complement(2529..3683)  
 /gene="yfgB"  
 /function="orf; Unknown function"  
 /note="Residues 1 to 384 of 384 are 99.47 pct identical to  
 residues 1 to 384 of 384 from Escherichia coli K-12 Strain  
 MG1655: B2517"  
 /codon\_start=1  
 /transl\_table=11  
 /product="orf, hypothetical protein"  
 /protein\_id="AG57627.1"  
 /db\_xref="GI:12516911"  
 /translation="MSEQLVTPENVYTKDGKINLIDLNRQQRFFKDLGEKPRADQ  
 VKRMVHYICDNFDEMTDINKVLRKLEVAEIRAEVEVEORSSDGTIKANINAGDO  
 RVEYVIEDRDATICVSSQVCALECFRCAGQGRNIRVSTICQVYRAATIVG  
 AKATGYGPIYVVMKGSEPLINNNVPAVEITLDDFGGLSKRRYTLSTGVVPA  
 LDKLGDMDVALAISLHAPNDEIRDEIYIPINKXKXEFLLAARLYLEKSNANGRV  
 IEYVMDLVHVNQSTEHQHLAELCKLPCKINLIPNPPAPAYGSSNSRDRFSKVL  
 MSYGVTTIVKRTKRGDDIDAGQLAGVDIDRTKRYLRKMGGEAIDIKAV"  
 complement(3833..4264)  
 /gene="ndk"  
 /note="synonym: 23781"  
 complement(3833..4264)  
 /gene="ndk"  
 /function="enzyme; Nucleotide biosynthesis: Purine  
 ribonucleotide biosynthesis"  
 /note="Residues 1 to 143 of 143 are 100.00 pct identical  
 to residues 1 to 143 of 143 from Escherichia coli K-12  
 Strain MG1655: B2518"  
 /codon\_start=1  
 /transl\_table=11  
 /product="nucleoside diphosphate kinase"  
 /protein\_id="AG57628.1"  
 /db\_xref="GI:12516912"  
 /translation="MAIRFRTSIKPNVAVKAVGNIFARPEAGKFTVQKMLHLY  
 EQARGFYAHGKPFPGDLVEFTMSGLIVSVLGEENAVQRHDLGATNPANLAGT

```

misc-feature      4397..9255
                  /note="O-island #106; Region of the EDL933 chromosome not
                  homologous to E. coli K-12 MG1655"
gene              complement(4405..4983)
                  /gene="23782"
                  /gene="23782"
                  /gene="23782"
                  /function="orf; Unknown function"
                  /note="No significant matches"
                  /codon_start=-1
                  /transl_table=-1
                  /protein_id="AA057629.1"
                  /db_xref="GI:12516913"
                  /translation="MLMHEKQGITATAIDLSQAKAMERALLINECLROYGEGSWY
                  SPRTSEVNSRRRLPHYRADYQTOIOPBLRLHILMPQISARPPMDPOKOLGA
                  CWRACEQVPSLNEGHLOINDALCNGQNCIACFOAMATEVELTIPAKIVNLHANK
                  VKCTCKSFLTFQONONCLYCGRHRYGMRTF"
CDS               complement(5259..6080)
                  /gene="23783"
                  /gene="23783"
                  /gene="23783"
                  /function="putative enzyme; Not classified"
                  /note="Residues 2 to 272 of 273 are 25.26 pct identical to
                  residues 6 to 276 of 286 from Genpept 118 :
                  g115002126|gb|AAD37317.1|AF135170.8 (AF135170) dimethyl
                  sulfoxide reductase subunit C [Yersinia pestis]"
                  /codon_start=-1
                  /transl_table=-1
                  /product="putative dimethyl sulfoxide reductase subunit C"
                  /protein_id="AA057630.1"
                  /db_xref="GI:12516914"
                  /translation="MHELPLMAFLILQASVGMTLWLAFLRHILITTSARPYWRPEL
                  LLAFAVGLISTLHLYGPLNMAHLSHSSWSLREITFGALYALGLSTLVA
                  LRRGGQMLMAAVGVIVDFCAQOYIHRSITTHQMYTLILGTGGLGSAVA
                  VFRSGILPQIDALRNGCVIVIALVILRLIYHOSITQIDLTANMOIATLPHAPLML
                  GQKPIITLSNGISVTCMMPFAVGGCKKNIPALFGSVMLVGSVMLEKRFPSIG"
gene              complement(6073..6510)
                  /gene="23784"
                  /gene="23784"
                  /function="putative enzyme; Not classified"
                  /note="Residues 4 to 134 of 145 are 61.83 pct identical to
                  residues 57 to 186 of 205 from Genpept 118 :
                  g111787122|gb|AAC73981.1|AE000191| anaerobic dimethyl
                  sulfoxide reductase subunit B [Escherichia coli]"
                  /codon_start=-1
                  /transl_table=-1
                  /product="putative dimethyl sulfoxide reductase subunit B"
                  /protein_id="AA057631.1"
                  /db_xref="GI:12516915"
                  /translation="MSHNVFAVYTLVSNCNHCADICTKNCPTMAMHRRPGDGIYRVNT
                  DKVGCYCAWSCPYGAPOMNEOTFGKMSKCDICDLOAKGPOPCVATCPGLGAIKFGP
                  IDELRKYGVSDVKGLPDSISITQPNLVIRKHAKEKASHHA"
CDS               complement(6699..9080)
                  /gene="23785"
                  /gene="23785"
                  /function="putative enzyme; Not classified"
                  /note="Residues 25 to 792 of 793 are 43.59 pct identical
                  to residues 3 to 784 of 785 from Genpept 118 :
                  g111787121|gb|AAC73960.1|AE000191| anaerobic dimethyl
                  sulfoxide reductase subunit A [Escherichia coli]"
                  /codon_start=-1
                  /transl_table=-1
                  /product="putative dimethyl sulfoxide reductase subunit A"
                  /protein_id="AA057632.1"
                  /db_xref="GI:12516916"
                  /translation="MDPDEKERITFPQVSRSRFLQATLITLPLFISSTAKAOSPA
                  SPEVTAPADKVPVPTGSCFDGKCDIRAHNRDGVVQITLIPNEDLPOMIRACV
                  RRGVRRKVVPRILKYMKRVKRGKEKFERISDEATLILANLKRITQOYPSAR
                  YVHGTATGCTGSDAMARLNLGTGYLEYVSHVSANTAAATPPTVGYAASGNSM
                  DTLDTKLIVILMGHNPETITGHTNYTFQKKKONGKTRIVYDPRKSDVSSLAOWKP

```

```

LPTDNLAMDAMMYIISNLHDKTFIDTFTLIGEDENSMRGPVANSILVAYLEGA
DGHTKPEWAEKITHVPAOSIROLADYATTPALIOGWGPHRICGERTARGSTLL
ASTGNVGIKGGMAGGYGSSNRKRGVDGMDENPOAKISIMNMQAADAQSKYTPQ
DKLKYDKRIDSINRIPLIFAGVIANONDVHOAKLLEDESKIEFIYSDJPMPSA
KTADVLIPETSEMERNITEPTWGTASYLLSKRIEPPEREDVDMLRDVAKTIGE
AEFSQGRDEKOWEHTIWEQTRLAMPDENIPDRATLOKTRHILFKSAPHILAFANRDP
QNNPEPTPSGKIEIFSKRLFDMQDEIPALSHVPAFEGPEKTLAKYPLQILTKGK
NRANSTQYANPWLQEVOTOKMLNPODAKORGISIBDSVKIYNDRGVSITPVEITPPI
US-09-921-992-78 (1-372) x AE005481 (1-11521)
Alignment Scores:
Pred. No.:      1 98e-134      Length:      11521
Score:          1866.00      Matches:      372
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels: 0
DB:             1             Gaps: 0
QY      1 MethisAsnGlnAlaProIleGlnArgGlySerThrArgIleTyrValGlyAsnVal 20
      |||||||
Db      1204 ATGATTAACCAAGCGCTCAATTCACAGTAAGTAACACGATTTTACGTGGGAATGAG 1145
QY      21 ProIleGlySpcGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
      |||||||
Db      1144 CCGATTCGCGATGCTGCTCCATCGCGGACGTACCTCCATACCAATACGCGTACGACAGAT 1085
QY      41 ValGlnAlaThrValAsnGlnIleTyrAlaLeuGlnArgValGlyAlaAspIleValArg 60
      |||||||
Db      1084 GTGGAACCAAGCGTCATCAATCAACAGCGCTGGAACGCGTGGCTGATATCGTCGCG 1025
QY      61 ValSerValProThrMetAspAlaAlaGlnAlaPheTyrSerIleTyrGlnGlnValAsn 80
      |||||||
Db      1024 GTTCCTGACCGACGAGTGAAGCGGGCGAAGCGTCAACTCAATCAACAGCGAGTATAC 965
QY      81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuTyrValAlaGluTyr 100
      |||||||
Db      964 GTGCGCGTGGTGGTGATCACTCCACTTCGATTCGCGTGAAGAGTACGGAATAC 905
QY      101 GlyValAspCysLeuAlaGlnIleAsnProGlyAsnIleGlyAsnGluGluArgMet 120
      |||||||
Db      904 GGCCTGATGCTGCTGCTATTAACCTGGCAATATCGTATGAAAGCGTATTCGATG 845
QY      121 ValValAspCysAlaAlaGlyAspAsnIleProIleArgIleGlyValAsnAlaGlySer 140
      |||||||
Db      844 GTGCTTACCTGTGCGCCGATAAACATTCCTCCATTCGCTATTCGCTTAACCGCGATCG 785
QY      141 LeuGluTyrAspLeuGlnGluTyrGlyGluProThrProGlnAlaLeuGluSer 160
      |||||||
Db      784 CTGGAATAAAGATCTGCAAGAAAGTATGCGAACCAGCGCGGCTTGCGATTC 725
QY      161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheTyrValSerVal 180
      |||||||
Db      724 GCCATGGCGCATGTGTATCATCTCGATCGCGCTGAATTCATCATGTTCAAGTACGCGTG 665
QY      181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaTyrGlnLeuAsp 200
      |||||||
Db      664 AAAGCGTGTGACGCTTCTCCGCTGCTGATCTTATCGTTGCGGCAAAACAATGAT 605
QY      201 GlnProLeuHisLeuGlyIleThrGluAlaGlyAlaArgSerGlyAlaValTyrSer 220
      |||||||
Db      604 CAGCGCGTGCATCTGGGGATACCGAAGCGGGTGGCGCGGACGCGGAGTAAATCC 545
QY      221 AlaIleGlyLeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
      |||||||
Db      544 GCCATGCTTATGATGCTGCTGCTGCTGCAAGGCAATCGGCACACGCGCGCATATCATG 485
QY      241 AlaAlaAspProValGluGluIleTyrValGlyPheAspIleLeuTyrSerLeuArgIle 260
      |||||||
Db      484 GCGGCGCATTCGCGTGAAGATAAAGATGCGTTCATATTTGAAATCGCTTCGATC 425
QY      261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
      |||||||

```



```

CDS      complement(2529..3683)
         /gene="yfgB"
         /function="orf, Unknown"
         /note="f384; 100 pct identical to yfgB_ECOLI SW: p36979"
         /codon_start=1
         /transl_table=1
         /product="orf, hypothetical protein"
         /protein_id="AAC75570.1"
         /db_xref="GI:1788865"
         /translation="MSEQLYTPENVTTKCKINLIDLRQQRREFKDLGKPPRAD
VKMWHYICDNFEDMDINKVLRKLEVEIKAVEVEEQRSSDGTIKALVGDQ
RVEFVYIPEDRALICVSSQVGCALCEKFCSTAOQGNRLNLYSEIIIGQVRAKIVG
AAKYGORPITNVVMGKGBELMLNNVPMAMEIMLDFGGLSKRRTVLTSGVPA
LDKGLMDVDAIATSLHAPNDEIRDEIYPIKKYNIETFLAARLYLEKSNAGRTV
IEVYMLDHVNDGTEHAHQIAELDKTPCKINLIPMNPPGAPYGRSSNSRIIDRSKVL
MSYFTTIVKTRGDDIDACGQLAGVIDTPTKRLKRRMGSEALDKAY"

promoter  complement(3778..3806)
         /note="factor Sigma70; predicted +1 start at 2642391"
         /product="factor Sigma70; predicted +1 start at 2642391"
         /note="factor Sigma70; predicted +1 start at 2642399"
         /gene="ndk"
         /note="synonym: b2518"
         /complement(3833..4264)
         /gene="ndk"
         /EC_number="2.7.4.6"
         /function="enzyme; Purine ribonucleotide biosynthesis"
         /note="f143; 100 pct identical to NDK_ECOLI SW: p24233"
         /codon_start=1
         /transl_table=1
         /product="nucleoside diphosphate kinase"
         /protein_id="AAC75571.1"
         /db_xref="GI:1788866"
         /translation="MATERPFSIIKPNVAVKNIENIFAREAGFKIVGKMLHTV
EQAGFPAHEHDKRPPDGLVEFMTSGPIYVSLSENAVQHRDLTGATNPANLAGT
LRADYADSLTNGTHGSDSVESAKEIAFFEGEVECPTR"

promoter  complement(4322..4350)
         /note="factor Sigma70; predicted +1 start at 2642935"
         /complement(4413..6725)
         /gene="pbpc"
         /note="synonym: b2519"
         /complement(4413..6725)
         /gene="pbpc"
         /function="putative enzyme; Not classified"
         /note="f770; This 770 aa ORF is 31 pct identical (43 gaps)
to 480 residues of an approx. 720 aa protein PBPF_BACSU
SM: p38050"
         /codon_start=1
         /transl_table=1
         /product="putative peptidoglycan enzyme"
         /protein_id="AAC75572.1"
         /db_xref="GI:1788867"
         /translation="MPRLTKRGCMITLAAPFLITLAKGADKILMPLEHVNARY
VVAODGTPMFPADADGITKRYPTTIEDVSPRYLEALTYEDRWTKHGVNPFYSARA
AMODLTSGRVISGSGSTLMQYARLLDPKPFEGKIRQLMWALDLEMLSEILITLY
LNRAPFGGTLOGIGASAQWALGKSPANISYSEAMALVLPAPSLRDPREIEAA
NKYLIERAAVGVMSREQVKESEREPITLARDROMQLAPLSRMLGSKSKDKITTL
DAGLORLEELIAONMKGRPLPRSSILAVVOTDKRWGVGSDVLDNDSRGHDVY
NSTRSGSVTKDPYVYGLADGLIHASLIDVPRRTDGPNGNDSPHGPISSEA
LVRSINLPAVOYLEAGKRPRAKIRNGTFLIPNGAAPNLILIGAGAKLEBMA
AYTAFARHGKAGKRLQDPDLERPLMSSAAAMTIRIMDEADPLPDSALPRAPL
AMKTGTSYGYDANAVIGNARYVIGITWRGDTGPVQGFAPASVPLLNQNNLTLS
RSANLPEDRPNSVYRGVYICWPGOSILEGCGNCRRLATWLIDSSQPTLLPEQEG
INGIRFPIIMDENGKRVADCPQAROEKINWPLPLEEWLPAESERRAARLPASTSCP
PYGDAOLPIOLGTGVGDAIITKRLPGAEEATLPLQSSGGAERWFWLNGEPLTERGRN
VTLLHTDKGDIQLLYMDVGOIATYKFWQ"

gene      complement(6726..11687)
         /gene="b2520"
         /complement(6726..11687)
         /gene="b2520"
         /function="orf, Unknown"
         /note="f1653; f1653"
         /codon_start=1

```

```

         /transl_table=1
         /product="orf, hypothetical protein"
         /protein_id="AAC75573.1"
         /db_xref="GI:1788868"
         /translation="MKRLRVACMLMALACDNNDNAPTAVKADPSEVTKAASSEN

Alignment Scores:
Pred. No.:      2,36e-134      Length:      13176
Score:          1866.00      Matches:      372
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      100.00%      Indels:      0
Db:              1      Gaps:      0

US-09-921-992-78 (1-372) x AE000338 (1-13176)

OY      1 MethiSaNgInAlaProIIeGInArGArGySSerThrArGIIeTyValGIyASyVal 20
         |||||||
Db      1204 ATGCATTAACCAAGGCTCCCAATTCACGTAGAAATCAACACGATATTACGTTGGCAATGTG 1145
         |||||||

OY      21 ProIIeGIyASpGIyAlaProIIeAlaValGInSerMetThrAsnThrArGThrThrAsp 40
         |||||||
Db      1144 CCGATTGGCGATGGTGGTCCCATCGCCCTACAGTCCATGACCAATACGCGTACGACAC 1085
         |||||||

OY      41 ValGIuAlarhValaSnGlnIleLysAlaLeuGluArGValGIyAlaAspIleValArg 60
         |||||||
Db      1084 GTCGAAGCAAGCGGTCAATCAATCAATCAAGCGCTGGAACCGGTGGCGGTGATATCGTCCGT 1025
         |||||||

OY      61 ValSerValProThrMetAspAlaAlaGluAlaPheLysIleuIleLysGInGluValasn 80
         |||||||
Db      1024 GTATCCGTACCGACATCGATGACGCGGACGAAGCGTTCAAATCATCAAAACAGAGGTATAC 965
         |||||||

OY      81 ValProLeuValAlaAspIIeHisPheAspTyraGIIeAlaLeuLysValAlaGluTy 100
         |||||||
Db      964 GTGCCGCTGTCGCTGACATCCATCGATCGATTCGATTCGCTGAAGTACGGGAAATAC 905
         |||||||

OY      101 GIyValaAspCysLeuArGIleAsnProGlyasnIIeGIyasnGluArGIleArgmet 120
         |||||||
Db      904 GCGCGCATGTCGTCGTATTAACCTGCAATATCGGTATGAAAGACGATTCGATTCGATG 845
         |||||||

OY      121 ValValaAspCysAlaArGAspLysAsnIleProIIeArGIleGIyValaSnAlaGlySer 140
         |||||||
Db      844 GTGGTGACTGGCGCGGATTAACATTCGATTCGATTCGATTCGATTCGATTCGATTCG 785
         |||||||

OY      141 LeuGIyLysAspLeuGInGluTySyrGIyGluProThrProGInAlaLeuLeuGluSer 160
         |||||||
Db      784 CTGGAAGAAAGATCTCAGAGAAAGATATGCCAGACCGCGACGCGTGGTGAATCT 725
         |||||||

OY      161 AlaMetArGHisValaAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
         |||||||
Db      724 GCCATGCGTCATGTTGATCATCTCGATCGCTGAACTTCGATCAAGTCAAGTCAAGTCA 665
         |||||||

OY      181 LysAlaSerAspValaPheLeuAlaValaGInSerTyraGLeuLeuAlaLysGInIleAsp 200
         |||||||
Db      664 AAAGCGTGTGACGTTCTTCCTGCTTGTAGTCTTATCTTTGCTTGCGCAAAACAGTCAAT 605
         |||||||

OY      201 GInProLeuHisLeuGIyIleThrGIuAlaGlyLeuAlaArgSerGIyAlaValaLysSer 220
         |||||||
Db      604 CAGCGTTGCATCTGGGATCACCAAGACCGGTGTGGCGACGGGGAGTAAATCC 545
         |||||||

OY      221 AlaIIeGIyLeuGIyLeuLeuLeuSerGIyIleGIyAspThrLeuArGIyValSerIeu 240
         |||||||
Db      544 GCCATTGTTTAAAGTCTCTGCTGCTCAAGGACACCGGACACCTCGGCTATCGCTG 485
         |||||||

OY      241 AlaAlaAspProValaGInGluIleLysValaGlyPheAspIleLeuLysSerLeuArGIle 260
         |||||||
Db      484 GCGGCGGATCCGCTGCAAGAGATCAAGTCCGTTTCATATTTCGAATCGTGGTATTC 425
         |||||||

OY      261 ArgSerArGIyIleAsnPheIleAlaCysProThrCysSerArgGInGluPheAspVal 280
         |||||||
Db      424 CGTTCGACGAGGATCACTTATCGCTGCCGACCTGCTGCGCAGAGAAATTTATGT 365
         |||||||

OY      281 IIeGIyThrValaSnAlaLeuGInArGLeuGluAspIleIleThrProMetAspVal 300
         |||||||

```

Db 364 ATCGGTACGGTTAACCGCTGGAGACAGCCGGAAGATATATACCTCCGATGACGTT 305  
 Oy 301 SerTlleIleGlyCyValValAlaSnnglyProGlyGluAlaLeuValSerThrLeuGlyVal 320  
 Db 304 TCGATTATCGCTGGCTGGTGAATGCCAGGTGAGCGCGCTGTTCTACACTCGGCGTC 245  
 Oy 321 ThrGlyGlyAsnLysSerGlyLeuTrpGlyGluAspGlyValArgLysAspArgLeuAsp 340  
 Db 244 ACCGGCGGCAACAGAAAACGGCGCTCATATGAGATGGCGTCCGCAAAAGCCGCTGGAC 185  
 Oy 341 AsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360  
 Db 184 AACCAAGATATGATGACACAGCTGCAACAGCAGCATTCGCGAAAGCAGTCAGCTGAC 125  
 Oy 361 GluAlaArgArgIleAspValGlnGlnValGlyLys 372  
 Db 124 GAAGCGCGTCGATGACGTTACGACGTTGAAAAA 89

RESULT 9  
 APO02561/c 296827 bp DNA linear BCT 24-APR-2003  
 LOCUS Escherichia coli O157:H7 DNA, complete genome, section 12/20.  
 DEFINITION APO02561 BA000007  
 ACCESSION APO02561.1 GI:13362592  
 VERSION  
 KEYWORDS  
 SOURCE Escherichia coli O157:H7  
 ORGANISM Escherichia coli O157:H7  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Escherichia.

REFERENCE  
 AUTHORS Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C. H., Kimura, S.,  
 Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T.,  
 Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T.,  
 Sasaki, C., and Shinagawa, H.  
 TITLE Complete nucleotide sequence of the prophage VT2-Sakai carrying the  
 verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7  
 derived from the Sakai outbreak  
 JOURNAL Genes Genet. Syst. 74 (5), 227-239 (1999)  
 MEDLINE 20198780  
 PUBMED 10734605  
 REFERENCE  
 AUTHORS Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M.,  
 Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and  
 Hayashi, T.  
 TITLE Comparative analysis of the whole set of rRNA operons between an  
 enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an  
 Escherichia coli K-12 strain MG1655  
 JOURNAL Syst. Appl. Microbiol. 23 (3), 315-324 (2000)  
 MEDLINE 20557356  
 PUBMED 11108008  
 REFERENCE  
 AUTHORS Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S.,  
 Yutsudo, C. H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T.,  
 Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasaki, C., and  
 Shinagawa, H.  
 TITLE Complete nucleotide sequence of the prophage VT1-Sakai carrying the  
 Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli  
 O157:H7 strain derived from the Sakai outbreak  
 JOURNAL Gene 258 (1-2), 127-139 (2000)  
 MEDLINE 20564182  
 PUBMED 11111050  
 REFERENCE  
 AUTHORS Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K.,  
 Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T.,  
 Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasaki, C.,  
 Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M., and  
 Shinagawa, H.  
 TITLE Complete genome sequence of enterohemorrhagic Escherichia coli  
 O157:H7 and genomic comparison with a laboratory strain K-12  
 JOURNAL DNA Res. 8 (1), 11-22 (2001)  
 MEDLINE 21156231  
 PUBMED 11258796

REFERENCE 5 (bases 1 to 296827)  
 AUTHORS Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and  
 Hayashi, T.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome  
 Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871,  
 Japan (E-mail:ken@gen-info.osaka-u.ac.jp,  
 URL:http://www.gen-info.osaka-u.ac.jp/,  
 Fax:81-6-6879-2047)  
 COMMENT genome project.

FEATURES  
 source  
 location/Qualifiers  
 1..296827  
 /organism="Escherichia coli O157:H7"  
 /mol\_type="genomic DNA"  
 /strain="O157:H7"  
 /sub\_strain="RIMD 0509952"  
 /db\_xref="taxon:83334"  
 414..1484  
 /gene="ECs3123"  
 414..1484  
 /gene="ECs3123"  
 /note="unknown"  
 /codon\_start=1  
 /evidence=not-experimental  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAB36546.1"  
 /db\_xref="GI:13362593"  
 /translation="MPTSIHLSRNIIDNNKLOPTEYIKESQOTAKLEPCDEFIKI  
 SOSRYGLSTSDSIANLAFETKASDASIDITOKIDVHSGFVRIRHVEVITYKN  
 EFADYEPTLTETGNEENEGTISVSEKGFISQRTININEETININEVITENI  
 HMTVOSPTEKYSIKINSKPYKTSKESIKSQLSDQKNITQYILTKKNTWKYA  
 SEERYDKRNSKSKETKNTYKOIVQKRLNOMKYINLSGLIDITDKKLAPLINDS  
 FSTYDFRFAISQKKNITFEDNSKFPVPEKRGIMVHSHNGOVLITGRYCEPGVATF  
 TQEQDKLIRYINIFTEPKLE"  
 complement(1708..2784)  
 /gene="ECs3124"  
 complement(1708..2784)  
 /gene="ECs3124"  
 complement(1708..2784)  
 /note="similar to GLPO\_ECOLI g11788572 percent identity  
 99 in 358 aa (Conserved in E.coli K-12)"  
 /codon\_start=1  
 /evidence=not-experimental  
 /transl\_table=11  
 /product="glycerophosphodiester phosphodiesterase"  
 /protein\_id="BAB36547.1"  
 /db\_xref="GI:13362594"  
 /translation="WKLTLKNSIAIMNSTIVMGSSANAAADSNKIVIAHNGASGYLP  
 EHTLPKAMAYAGADYLEDLVMTKDHVLVLDHYLDRTVDVADRPDPAKRGYR  
 YAIPTLDEIKSLKFTGEGDIENGKKVOTYGRFPMKSGDFRVHFFEEIEFVGLNH  
 STGKNIGIYPEIKAPWPFHOGKIDIAKTLVIRKYGVTGKDDVYVLOCPPADELKR  
 KNELEPKMGMDNLVOLIAVYDMNETOKODSGSVNRYNPMKPGKMKVAYEADG  
 IGPVHMLIEFTSOPGKIKLNGVODAKNLVHPITVRSDKLPEYTTDVNQLYDAL  
 YNKAGVNLFTDPKAVKFLNKE"  
 complement(2789..4147)  
 /gene="ECs3125"  
 complement(2789..4147)  
 /gene="ECs3125"  
 complement(2789..4147)  
 /note="similar to GLPT\_ECOLI g11788573 percent identity  
 99 in 452 aa (Conserved in E.coli K-12)"  
 /codon\_start=1  
 /evidence=not-experimental  
 /transl\_table=11  
 /product="sn-glycerol-3-phosphate permease"  
 /protein\_id="BAB36548.1"  
 /db\_xref="GI:13362595"  
 /translation="MLSIFRAPKALPAELIDPYTRLRMQILGIFEGYAYVLY  
 RKNRLAPLTVLEGEFSRGGDFALSDISLAYGSKTMSVSRSRNPRLFAGLIL  
 AAAYVLFMGFVPMTSSIAVWFVILFLCGFGQGMPCRTWVHMSKRGKGIYSV  
 WNCANHNGGGLPPLFLFLGMAFMFDMALAYMPAFCAIIVAFAPFAMRPTPOSGIP  
 PTEYKNDYPDYNEKAEQELTAQIEMOVVLVLPKILMYAIANVPIYLIRYGLDMS  
 PTYLEKVKHFALDKSSNAFFPYAGIPGLIGMSDKYFRNGRATGVFEMTLVTI  
 ALIYVWNPAGNPVDMICMIVIGFLYIGFVMLIGLHLELAPKRAAGTAAGPTGLG







[illegible]

/protein\_id="AA081323.1"  
/db\_xref="GI:26109121"  
/translation="MEFNSLFMVSPILGLVFEVAMLAGETIDSIAGGGGLTTPALM  
AAGSPANALATNKLQACGSGISATITFIRKYSISDOKNTAMTFVSGMGLIYO  
YVADVRLQILPILVICITGLFLLMPLGEGDQRRRTGIPPLLIAGCGYFDDGFCG  
PAAGSFYALAEVLTCGFNLAKATNAHALNATNIGLLIFLIGKVIATGFMALVG  
QFLGARMSRLVSLCGKILRPMIVIVSAVMSAKLXDSHGEILHMLGN"  
/complement(5338..6162)  
/gene="mepa"  
/locus\_tag="c2874"  
/complement(5338..6162)  
/gene="mepa"  
/locus\_tag="c2874"  
/EC\_number="3.4.99.-"  
/function="enzyme: Murein sacculus, peptidoglycan"  
/note="Escherichia coli K-12 ortholog: b2328; Escherichia  
coli O157:H7 ortholog: z3591"  
/codon\_start=1  
/transl\_table=11  
/product="Penicillin-insensitive murein endopeptidase  
precursor"  
/protein\_id="AA081324.1"  
/db\_xref="GI:26109122"  
/translation="MNKTAIALALILASASLAAATPMOKITOPVPGSAOSIGSFNSGC  
ITGADTLPJOSEHYOVMTDQRRYEGHDMYFIORLSROYSNIGMGTIVYALKMGA  
GGRNGASHQTCGLDVIDPILPKTMTSAQLBPDALVSRDGSKVVYALKMPET  
FSLKLAQKQDVTRITFNPALIKOOLCIDAQTDPRDLAKYPMFOHRAHMHVRLRCPA  
DSLCEDDPLPPDGCAGELQSMFEPEPKPTTPEKKTPTPLPSCALIDENVT"  
/complement(6166..7254)  
/gene="aroc"  
/locus\_tag="c2875"  
/complement(6166..7254)  
/gene="aroc"  
/locus\_tag="c2875"  
/EC\_number="4.2.3.5"  
/function="enzyme: Amino acid biosynthesis: Chorismate"  
/note="Escherichia coli K-12 ortholog: b2329; Escherichia  
coli O157:H7 ortholog: z3592"  
/codon\_start=1  
/transl\_table=11  
/product="Chorismate synthase"  
/protein\_id="AA081325.1"  
/db\_xref="GI:26109123"  
/translation="MAGNTIGOLFRRVTFEGSHGLAGCIVDVPQIPLTEADLOH  
DIDRRPGTSRTTORREPDQVILSGVEGVTGTSIGLIENTDOSODYSALQV  
FRHGRADYEQKYLGRDRGSGSSARETAMRAAALAKYLAELKSGIELRGCLTQ  
MGDILLETKMSOVONPEFCPDPKIDALMRALKKEGSDISAKTYVAVGVPAG  
ICEPVFDRDLADIAHALMSINAKVEIGEDPDYVALRGSORDEITTDGFGSNHAG  
ILGSISSGQIITAHMAKPTSSITTPGRTINNFGEVEKITGRHDPVCGRAPVIAE  
AMLAIVLMDHLRQKQNDVKTDIRPW"  
/complement(7286..8551)  
/gene="yfcB"  
/locus\_tag="c2876"  
/complement(7286..8551)  
/gene="yfcB"  
/locus\_tag="c2876"  
/EC\_number="2.1.1.72"  
/function="putative"  
/note="Escherichia coli K-12 ortholog: b2330; Escherichia  
coli O157:H7 ortholog: z3593"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical adenine-specific methylase yfcB"  
/protein\_id="AA081326.1"  
/db\_xref="GI:26109124"  
/translation="MAGIGFLNQLPWFADFTLRSDNRYLNLASPRHOLAKOR  
IVLAKCCFLHLSFSISNPEVYPTRTSIDRKILFSPSPRNADLSSSMQTSRN  
YCHALAEVYDKIFVDEAVNELOTIDMRWSYRSANINWGHGTDMPDMAVOLV  
LPSLYLPDIPEDMTRARLISSEKRIYERIRVNERIPVATITRAREFCGHEEVVD  
ERVLVPRSPIGELLINKFAGLISKPOHILDMCTGSCIALACATAPPAEDAVDIS  
PDALAAEONIEEHGLIHNVILIRSDLFRLPVQYDLVITNPVYDADMSDMLPEY  
RHEPELGASGTGDKLIRITLGNADYLAADGVGLICEVNSMVHLMBOYPDVPTWL

Alignment Scores:  
Pred. No.: 1,47e-132 Length: 300099  
Score: 372 Matches: 372  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0  
US-09-921-992-78 (1-372) x AE016764 (1-300099)  
QY 1 Methiasnnglnalabproilleginararglyssertharqleqyvalaqlasnal 20  
DB 188410 ATGCATACACGAGCTCAATTCACCTAGAAATCAACGATATTACCTGGGAAATGTC 188351  
QY 21 Proilleglyaspcllylabproillealavallglnsermethrasnthrarqthrnaap 40  
DB 188350 CCATTTGGCGATGCTGCTCCATTCGCCGGRACGTCCATGACCAATCGGTTGACAGAC 188291  
QY 41 Valglualathrvalasnnglnilleysalaleugluargvalaqlaaspillevalarg 60  
DB 188290 GTGGAACCAACGGTCATCAATCAACGAGCGCTGGAACGCTTGCCCTATATCCGTCGT 188231  
QY 61 Valservalprothmethaspaialaialaaphelystreulleysglnvalaasn 80  
DB 188230 GTCTCCGTACCGACGATGACGCGGCGAGAACGCTTCAATCATCAACGAGCTTAC 188171  
QY 81 Valproleuvalalaaspillehispheasprytrargillealaleuysvalaialguty 100  
DB 188170 GTGCGCGTGGGGCTGACATCCACTTCGACTATCCGCTTGGTGAAGTGAAGTGAATGC 188111  
QY 101 Glyvalaspcysleuaurgilleasnprogllyasnilleglvasnugluarqileargmet 120  
DB 188110 GGCCTGATGCTGCTGCTATTAACCTGGCAATATGCGTATATGAAGCGCTATTCGATG 188051  
QY 121 Valvalaspcysalaargaspilysasnleproilleargilleglvalaasnlaaglyser 140  
DB 188050 GTGTTGACGTGTGGCGCGAATAAATCATCCGATTCGATTTGGCGCTTAATGCCGATGC 187991  
QY 141 Leugluysaspleuglngluysgylglnprothrproglinalaleuugluser 160  
DB 187990 TTGGAAAAAGATCTCGCAAGAAAGATATGCGCAACGCGCGCGGCTTGCTGATCT 187931  
QY 161 Alamelarghisvalasphasileuaspargleuasnphaspglnphelysvalserval 180  
DB 187930 GCATATGCTCATGTGATGATCATCTGATCGCTGAATCTCGATCAAGTACAGCGT 187871  
QY 181 Lysalaseraspvalpheelualavajluserytrargleuualaleuuglnilasp 200  
DB 187870 AAAGCGTGTGACGCTCTCTGCTGCTGAGCTTATGCTTTGCGCAAAACAATCGAT 187811  
QY 201 Glnproleuhsleuglyilethrqlualaglylaalarsergilaalavallysser 220  
DB 187810 CAGCCGTCGATCTGGGATGATCAGCAACGCGGTGGCGCTACGCGGCGAGTAAATCC 187751  
QY 221 Alalileglyleuglyleuleuleusergluylileglasprhrlleuargvalserleu 240  
DB 187750 GCTATTGGTTAGCTGCTGCTGCTGCTGAAGCAACGCGCGAGCTGCGGTATCGCTG 187691  
QY 241 Alaalasprovalglnuglnleuleysvalgylpheaspileleuysserleuarglle 260  
DB 187690 GCGGCGGATCGGTGGAAGATCAAGTCCGTTTCGATTTTGAATTCCTGCGCATC 187631  
QY 261 Argserargglyileasnphellealacysprothrctysseraraglnuglnpheasval 280  
DB 187630 CATTTCGGCGGATGATCACTTATCCCGGACCTGCTGCGCTGCGCAAGATTGATGTT 187571  
QY 281 Ileglythrvalaasnaleuugluarqleugluaspillelethrprometaspaal 300  
DB 187570 ATCGGATCGGTTAAACGCGCTGAGCAACGCGTGAAGATATCATCTCCGATGACGTT 187511  
QY 301 Serilleileglycyavavalaasnuglyproglylualaleuvalserthrleuglyval 320  
DB 187510 TCGATTATCGGCTGCTGCTGTAATGCGCCAGGTGAGCGCGCTTTCACACTGCGGCTC 187451

QY 321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340  
 Db 187450 ACTGGCGCAACAGAAAGAGGGCTCTATGAGACATGCGCTGCCAAGACCGTCTGGAC 187391  
 QY 341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360  
 Db 187390 AACACGATATGATCGATCGATCGAGAACCGCATTCGTGCGAAGCAGCTGAGTGCAC 187331  
 QY 361 GluAlaArgArgIleAspValGlnGlnValGluLys 372  
 Db 187330 GAAGCGCTCGAATGACGTTCAAGCAGGTTGAAAA 187295

RESULT 11  
 AE015271/c 10225 bp DNA linear BCT 18-OCT-2002  
 LOCUS Shigella flexneri 2a str. 301 section 234 of 412 of the complete genome.  
 ACCESSION AE015271 AE005674  
 VERSION AE015271.1 GI:24052928  
 KEYWORDS Shigella flexneri 2a str. 301  
 SOURCE Bacteria: Proteobacteria: Gammaproteobacteria: Enterobacteriales: Enterobacteriaceae: Shigella.  
 ORGANISM Shigella flexneri 2a str. 301

REFERENCE 1 (bases 1 to 10225)  
 JIN, Q., YUAN, Z. H., XU, J. G., WANG, Y., SHEN, Y., LU, W. C., WANG, J. H., LIU, H., YANG, J., YANG, F., QU, D., ZHANG, X. B., ZHANG, J. Y., YANG, G. W., WU, H. T., DONG, J., SUN, L. L., XUE, Y., ZHAO, A. L., GAO, Y. S., ZHU, J. P., KAN, B., CHEN, S. X., YAO, Z. J., HE, B. K., CHEN, R. S., MA, D. L., QIANG, B. Q., WEN, Y. M., HOU, Y. D. and YU, J. D.  
 Genome sequence of Shigella flexneri 2a: insights into pathogenicity through comparison with genomes of Escherichia coli K12 and O157  
 Nucleic Acids Res. 30 (20), 4432-4441 (2002)

TITLE  
 JOURNAL  
 PubMed  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry of Public Health, 100 Yangxin Jie, Xuanwu Qu, Beijing 100052, P.R. China

FEATURES  
 source  
 Location/Qualifiers  
 1..10225  
 /organism="Shigella flexneri 2a str. 301"  
 /mol\_type="genomic DNA"  
 /strain="301"  
 /serotype="2a"  
 /db\_xref="taxon:198214"  
 148..1440  
 /gene="xseA"  
 /locus\_tag="SF2555"  
 148..1440  
 /gene="xseA"  
 /locus\_tag="SF2555"  
 /note="Residues 1 to 430 of 430 are 97 pct identical to residues 27 to 456 of a 456 aa protein from Escherichia coli K12 ref: NP\_417004.1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="exonuclease VII, large subunit"  
 /protein\_id="AAN44055.1"  
 /db\_xref="GI:24052928"  
 /translacion="MGQVWISGEISNFTOPASGHMYFLTKDDTAQVRCAMFRNSRRVTFPQHQOQVLRANITLPEPRGYQITVESMOPAGEGLLOQKEOLKAKLOAECLDQYKRPPLSPAHCVGVTSTKALHDLHLKLRDPSLPYIITPSVOCDDAAGQIVRAELANRNECDVLIVRGSGLEDLMSFNDERVARITASRPVYSAYGHEVDITADFAVDRAPLPSAAEVVSRNOCELRLROVSARORLEAMADYILANTRKTRFTQIHRLLOOHQRLRLAROOTMHERQKRSPFLNOLKRAQGGQORLTORLNOQNPORIKHRTOTRIQOLEYRLAEITLRAQLSATRETFRGNVAVTHLEAVSLTLARGYSVTATGDNVVK

gene  
 complement(1437..1688)  
 /gene="yfgJ"  
 /locus\_tag="SF2556"  
 complement(1437..1688)  
 /gene="yfgJ"  
 /locus\_tag="SF2556"  
 /note="Residues 1 to 83 of 83 are 97 pct identical to residues 1 to 83 of a 83 aa protein from Escherichia coli K12 ref: NP\_417005.1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="orf, conserved hypothetical protein"  
 /protein\_id="AAN44056.1"  
 /db\_xref="GI:24052930"  
 /translacion="MNVESMATGATGHMELHCPKCOHVLDDQDGHARCPSCGFTEKMA LCPDCHOPLOVLRKACGADVDFCOHGHGLSKRVEFLVA"  
 complement(1721..3232)  
 /gene="yfgK"  
 /locus\_tag="SF2557"  
 complement(1721..3232)  
 /gene="yfgK"  
 /locus\_tag="SF2557"  
 /note="Residues 1 to 503 of 503 are 99 pct identical to residues 1 to 503 of a 503 aa protein from Escherichia coli O157:H7 EDL933 ref: NP\_289064.1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative GTP-binding factor"  
 /protein\_id="AAN44057.1"  
 /db\_xref="GI:24052931"  
 /translacion="MRCIMTYKNEALNVPVVALYGRPNVSKSTLFENLRTDALVADPQITRKRKIGREIEFTICDTGIDTGVETRAEQLSLAEADVLFLV DARGLMPADALIAIKHLSRKPFLVVKNDGDDPDQAVDFSLGGEIYPIASH GRGVLSLIEHVLPMMEDIAEQEEDAEYMAQFEAEENGEEDDFDQSPILK AIVGPNVGNKSTINRILIGERVVYDMGTTROSIYIPMRDGRVYLITPACVRR KRIDAYKVFSTVITKILIEDANVVMVTDAREGISODLSLGFILNSGSLYIVN GMDISQVKEQVQVETEDRCPIDFPAVHFISSLHSGVGNLESYREAVDSTRV GSWLCTRMTAVNAVEDHOPPLVGRVRLKTAHAGGVNPVIVIHGNQVKDLPDSYKRY LMYFRSLDWSGSPTRIQEKGENPANKRNTLTPQMRKRKLKMKMK"  
 complement(3311..4489)  
 /gene="yfgL"  
 /locus\_tag="SF2558"  
 complement(3311..4489)  
 /gene="yfgL"  
 /locus\_tag="SF2558"  
 /note="Residues 1 to 392 of 392 are 99 pct identical to residues 1 to 392 of a 392 aa protein from Escherichia coli K12 ref: NP\_417007.1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative dehydrogenase"  
 /protein\_id="AAN44058.1"  
 /db\_xref="GI:24052932"  
 /translacion="MQLRLKLLPGLSVTLGSCSLFNSSEEDVYKNSPLPTVENOFTPTWMSVSGSIGNFIENLHPALADNVVYADRAGLKALNADDGKRTWSIAEEDG WFSKEPALISGVVSGHYIKKIVNOMSLSRGSAPATFCAVAGGDSGRVSA VIMEQGMIMQOIRISQATGSTEDRSLSDVITFVVVNGVAFALVGNLHRLITSPVLYN IMKRELGSVNDFTVDSNRLYLDQNDRVALLTIDGGVTLMAQSDHLHRLITSPVLYN GNLVYDSSEGYLHMVINEDEYRFAVQKVDSSGFTPEVADGKLLIQAKDGIVYSYTR

gene  
 complement(4500..5120)  
 /gene="yfgM"  
 /locus\_tag="SF2559"  
 complement(4500..5120)  
 /gene="yfgM"  
 /locus\_tag="SF2559"  
 /note="Residues 1 to 206 of 206 are 98 pct identical to residues 1 to 206 of a 206 aa protein from Escherichia coli K12 ref: NP\_417008.1"  
 /codon\_start=1  
 /transl\_table=11

```

/product="orf, conserved hypothetical protein"
/protein_id="A044059.1"
/db_xref="GI:24052933"
/translation="MEYENDQVAVKRFPAENKALAVGILGVGALGWRMNS
HWDASARSASIAVONAVTAVSEKPDSPAPAEAKENKTYGALSLAQOFVKN
ELEKAAALQDGLADTSDETKAVINRIARVGVOKADAAKTKTIDTIKGEWMAIV
ADLRGEALSKGQKQKARSAWENGVSDVTALSEMMQKININLSI"
/gene="hisS"
/complement(5138..6412)
/locus_tag="SF2560"
/gene="hisS"
/locus_tag="SF2560"
/notes="Residues 1 to 424 of 424 are 99 pct identical to
residues 1 to 424 of a 424 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_289067.1"
/codon_start=1
/transl_table=1
/product="histidine tRNA synthetase"
/protein_id="A044060.1"
/db_xref="GI:24052934"
/translation="MAKNIDAIKGMNDYLGEETAIWRIESTLKNVIGSYGSRIRP
IVEQIPLFKRAIGETDVVEKEMTFEDRNGDSLTLRPEGAGCVRAGIEHLIYNE
QRLWYIGPMERHERPQKGRYKQFHLQCEVFGLOGPIDAELMLTARWRAIGIEH
VTLLENISGLEAANYRADALVAFLEQHEKIDEDCKRRMTNPLRVLDSENPEVAL
LNDAPAGDYDEESREHFNAGLCKLESAGIAYVNOIRLVGLDYNRVTEVNTSL
GSQCTVACAGRYDGLVEQIGRATPAVGFMGLERVLVAVNPERKADPVYDIYV
ASGADTOSAMAAALAEERLDELGTAKIMTHNGGKPKFQFADADKGAARVAVLGESEV
ANGTAVAKDRSGEQTAVADSVAAHLRTLLG"
/complement(6515..7642)
/gene="gcpE"
/locus_tag="SF2561"
/complement(6515..7642)
/gene="gcpE"
/locus_tag="SF2561"
/notes="Residues 1 to 372 of 375 are 99 pct identical to
residues 1 to 372 of a 372 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_289068.1"
/codon_start=1
/transl_table=1
/product="orf, conserved hypothetical protein"
/protein_id="A044061.1"
/db_xref="GI:24052935"
/translation="MHNQAPIQRRKSTRIVGNVPIGDCAPAVQSMNTNRTTVEAT
VNOIKALRFGADIVRVSPPTDAEAKRLKQVNAVSLVADIHFDYRIALKVAEYV
DCLRNPGIGNEERIRVAVDCARDKNIPIRIGVAGSLEDLQEKGEPPQALLES
AMRYDHLDRNFDOFKYVKASDVLEVESRLAKIIDPLHIGIEAGARGAV
KSAIGLILSGISGIDTILRVSLADPVEIIVGPDILSLIRSRGINFACPGCSRO
EFDVIGTVALSKQRLLEDITTPMDVSITICVYVNGPEALYSTLCYTGKNGKSGLYEDGV
RKDRIDNMDIDQLERIRAKASOLDEARIDVOVGKLT"
/complement(7669..8682)
/gene="yfgA"
/locus_tag="SF2562"
/complement(7669..8682)
/gene="yfgA"
/locus_tag="SF2562"
/notes="Residues 1 to 337 of 337 are 99 pct identical to
residues 1 to 337 of a 337 aa protein from Escherichia
coli O157:H7 ref: NP_311405.1"
/codon_start=1
/transl_table=1
/product="putative membrane protein"
/protein_id="A044062.1"
/db_xref="GI:24052936"
/translation="MNTFATHDQNEALTTGARLRNAREQGLSQOAVARLCKIVSTV
RIERDKARADLASTFTLVYLVFVYIGLSGAMWODHRAQOEITTMDOSSAELSNSEO
GRRKRKRDCHLMTFTTLVYLVFVYIGLSGAMWODHRAQOEITTMDOSSAELSNSEO
GOSVPLNTSTTDPATSTPPASVDTRANTOTPAVTAAPVADPQNAVPSQNAV
DTAAATPPTAATTPDGAAPLPIDQACVTPPADPNALVNNFTADCLVETDITGKLF

```

```

Percent Similarity: 99.73%
Best Local Similarity: 99.46%
Query Match: 99.41%
DB: 1
Gaps: 0
US-09-921-992-78 (1-372) x AE015271 (1-10225)

```

```

1 MetHisASnglnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlnAsnVal 20
DB ATCATACACAGGCTCCATTCACGTACGAAATCAACAGTATTACCTGGCAATGTG 7583
QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
DB 7582 CCGATTGGCGATGCTGCTCCACAGCGGACATCCATGCAATACGGCTACGACAGT 7523
QY 41 ValGlnAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
DB 7522 GTCCAGCAACGGCTCAATCAACAGCGCTGGAACGGCTGCGCTGATTCGCGCT 7463
QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnIleValAsn 80
DB 7462 GTCTCCGTACCGACGATGAGCGGCGAGAGCGCTTCAAACTCAACAGCGGATTAC 7403
QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
DB 7402 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCTGGAAGTACGGGAATAC 7343
QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
DB 7342 GCGCTCATGTTGCTGCTATTAACCCGTAAATATCGTAAATACAGCTATTCGATC 7283
QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
DB 7282 GTGGTTACCTGTGGCGCGATAAACATTCGCCATTCGCTATTCGCTTACGGCGGATCG 7223
QY 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuGluGluSer 160
DB 7222 CTGGAAGAAAGATCTGCAGAAAGATATGCGAACCGCGCGGCTGTGGATTC 7163
QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnHisPheAspGlnPheLysValSerVal 180
DB 7162 GCCATGCGCTCATGTGATCATCTGATCGCTGAACTTCGATTCGATTCGATTCGATTC 7103
QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
DB 7102 AAAGCGTGTGACGCTTCCTCGCTGCTGATCTATGTTGCGGCAAAACAGATCAT 7043
QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyAlaArgSerGlyAlaValLysSer 220
DB 7042 CAGCCGCTGCATCTGGGGATCACCGAACCCGCTGCTGACCGACCGGGCGATTAATCC 6983
QY 221 AlaIleGlyLeuGlyLeuLeuLeuSerGlnIleGlyAspThrLeuArgValSerLeu 240
DB 6982 GCCATTGGTTAGGCTGCTGCTGCGGAGGCAATCGCGACACGCTCGGTATTCCTG 6923
QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
DB 6922 GCGCGCATTCGCGTGCAGAGATCAAACTCGGTTTCGATTTTGAATTCCTCGATTC 6863
QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
DB 6862 CGTTCGACAGGATTAACCTTATCCGCTGCCGACCTTTCGCGTCAAGAAATTTGATGT 6803
QY 281 IleGlyThrValAsnAlaLeuGlnGluIleArgLeuGluAspIleIleThrProMetAspVal 300
DB 6802 ATCGGTACGGTTAAACGCTGAGAGACGCTCGAAGATATCATCATCCGATGACGCTT 6743
QY 301 SerIleIleGlyCysValValAsnGlyProGlyGlnAlaLeuValSerThrLeuGlyVal 320
DB 6742 TCGATTATCGCGTGGTGTGAATGCCAGGTGAGCGGCTTTCATCACTCGGCGCTC 6683
QY 321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340

```

Alignment Scores:  
Pred. No.:  
Score:

1 2e-133 Length: 10225  
1855.00 Matches: 370



```

/transl_table=11
/product="ISSf14orf"
/protein_id="AaP17892.1"
/db_xref="GI:30042167"
/translation="MTESDSESVQFVIGVGDKDRHHAVAINRSGRKLFKALPNDE
NKRSLISDKHGOILLVYDQATIGALPAVARSGVAVGLPGIAMRIRIDLNG
EATDVRDAIIEAARTPLHALRTKLAEQIAELSMICGDDDLAQAOTQASNRK
GLTQIHAPERVLGPRLEHAPVILDLORPSPERKLAGSKELAQOLCLAPRIGR
LAADIAQALAEQTVVVGTAANAAYVLPRLAQLITLAKQDEVALVEQPRVLAHPIK
VLTSMQGVRTARLRLTEVACRFASAHLAAYAGIAPYTRRSQSSITREHSHRRN
KALKRALFLSAFALRDLPSRAYVTRKMSGKRHNQALIALARRKCVLFAAMMDGTF
YTPQGS"
/gene=".8442
/gene="ssca"
/locus_tag="S2740"
/notes="truncated"
/pseudo
complement(9261..10037)
/gene="sscb"
/locus_tag="S2741"
complement(9261..10037)
/gene="sscb"
/locus_tag="S2741"
/function="phenotype: Not classified"
/notes="residues 1 to 258 of 258 are 94.96 pct identical to
residues 4 to 261 of 261 from Escherichia coli K-12 :
B2522"
/codon_start=1
/transl_table=11
/product="enhancer of serine sensitivity"
/protein_id="AaP17893.1"
/db_xref="GI:30042168"
/translation="MSETKNEEDLLEKATPEAPHPAPFRTLESTVWVPGTAQGE
AYVDSLADLQHEKEDGTSVPEPTSLAEQAVDEDAFVWPVATLEMTLGLT
FLNKLPTGKEFMRETSILGERGNLSCOEILEGSESLISVAPPPQMDISLT
LFTKIKPKRAFLCSIKENEAPNLLIGLEADDIEITLQATSVATDPLPDEPID
ICQVKKGEKGSFHTETIHAFYERRMGGELRDKQRNRII"
10137..10170
/notes="REP (repetitive extragenic palindromic) element,
contains 1 REP sequence"
complement(10179..11462)
/gene="pepb"
/locus_tag="S2742"
complement(10179..11462)
/gene="pepb"
/locus_tag="S2742"
/function="putative enzyme: Not classified"
/notes="residues 1 to 427 of 427 are 92.50 pct identical to
residues 30 to 456 of 456 from Escherichia coli K-12 :
B2523"
/codon_start=1
/transl_table=11
/product="putative peptidase"
/protein_id="AaP17894.1"
/db_xref="GI:30042169"
/translation="MTBAMKTTTSTOPADARWGEKATYSINNDGITHLNGADLGLI
QRAARKIDGLGKIKHVLSEGMADRCWAFWQYKGRGTRKVMPLDADQROELN
RLMILDMWRDITINAPAEELGSQLAORAVDLISNAGDRVYRTIKGEDLAEQGMGL
HTVGRSESRPVLLADYNTPGKEAPVYACLVGKSTPDSGGYSITKQNGKVMKSD
MGAAATYTGALAAFAIMRGINKRKLELCCADNLLSGNAFRIGDITITRNKAYVMAT
DAEGRVYADGLIDASAKRPMITDAATLTGAAKTALGANDYHALFSPDDALAGLLAS
ASQNEPFWRLPLAEFHRSQLPSNFAELNNTGSAAYPRAGASTAGFUSHFEVNTQOGM
LHIDCSATYRKRAPEQMSAGATGLGVRTIANLTA"
complement(11521..11721)
/gene="yfhj"
/locus_tag="S2743"

```

```

DB: 1 Gaps: 0
US-09-921-992-78 (1-372) x AE016987 (1-290380)
QY 1 MethisAsnGlnAlaProIleGlnArgGlySerThrArgIleTyrValGlyAsnVal 20
1209 ATGATATACCAAGCGCTCAATTCAACGTGAATAATCAACAGTATTACGTGGGAATGTG 1150
QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
1149 CCGATTGGCCATGATGCTGCCATCCGCTGACGTGACATGACCAATACGCTAGACAGAT 1090
QY 41 ValGlnAlaThrValAsnGlnIleGlyAlaLeuGlnArgValGlyAlaAspIleValArg 60
1089 GTCCAAACCAACGCGTCATCAATCAAGCCCTGGAAAGCGCTGTGATTCACGTCCTG 1030
QY 61 ValSerValProThrMetAspAlaAlaGlnAlaPheGlySerIleGlyGlnGlnValAsn 80
1029 GTCTCCGTACCGAGATGACGCGGCGAAGCGTTCAACTCATCAACACGCGGTTAC 970
QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuGlyValAlaGluTyr 100
969 GTGCGCGTGGTGGTGACATCCACTTGCATTCGCGTGGTGGTGAAGAGGAAATAC 910
QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGlnGlnArgMet 120
909 GCGCTGATTTGCTGCTATTAACCCGTAATATCCGTAATGAAGACGCTATTCGATG 850
QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
849 GTGGTTACGTGGCGCGAATAAACAATTCGATTCGCTATTGGCGTTAACCGCGATCG 790
QY 141 LeuGlnLysAspLeuGlnGlnLysTyrGlyGluProThrProGlnAlaLeuGluSer 160
789 CTGCAAAAAGATCTCGCAAGAAAGATATGCGAACCAGCGCGCGTGTGGAATCC 730
QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnHisAspGlnPheLysValSerVal 180
729 GCCATGTGCTCATGTTGATCATCTGATCGCTGAACCTTCATCTCAAGTACAGCGTG 670
QY 181 LysAlaSerAspValPheLeuAlaValGlnSerTyrArgLeuLeuAlaLysGlnIleAsp 200
669 AAGCGCTGACGCTTCTCTCGTGTGAGCTTATGTTGCGGCAAAAACATCGAT 610
QY 201 GlnProLeuHisLeuGlnIleThrGlnAlaGlyAlaArgSerGlyAlaValLysSer 220
609 CAGCGCTGATCTGGGATCACCCGAAGCCGTTGTCACGACGCGGCGAGTAAATCC 550
QY 221 AlaIleGlyLeuGlyLeuLeuLeuSerGlnGlyIleGlyAspThrLeuArgValSerLeu 240
549 GCCATGTGTTAGTGTGCTGCTGCTGCGAAGGATCGGCGACACGCTGCTGATCGCTG 490
QY 241 AlaAlaAspProValGlnGlnIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
489 GCGGCGCATCGCGTGAAGATCAAAAGTGTTCATATTGTAATGCTGCTGATC 430
QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGlnPheAspVal 280
429 CGTTCGGAGAGGATTAATCTCATGCTGCCGCGACGCTGCTGCGTGAAGATTGAGCTT 370
QY 281 IleGlyThrValAsnAlaLeuGlnIleArgLeuGlnAspIleIleThrPrometAspVal 300
369 ATCGGTACGGTTAACCGCGCTGACACACGCTGGAAGATATCATCATCTCCATGAGAGTT 310
QY 301 SerIleIleGlyCysValAlaAsnGlyProGlyGlnAlaLeuValSerThrLeuGlyVal 320
309 TCGATTATCGGCTGCTGCTGATGATGCGCCAGGTGAGCGGCTGCTTACACTCGGCGTC 250
QY 321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGlnAspGlyValArgLysAspArgLeuAsp 340
249 ACCGCGCGCAACAAGAAAAGCGGCTCTATGAATGCGCGGCAAGACGCTGAC 190
QY 341 AsnAsnAspMetIleAspGlnLeuGlnAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360

```

## Alignment Scores:

```

Pred. No.: 1e-131 Length: 290380
Score: 1855.00 Matches: 370
Percent Similarity: 99.73% Conservative: 1
Best Local Similarity: 99.46% Mismatches: 1
Query Match: 99.41% Indels: 0

```





```

/db_xref="GI:16421061"
/translation="MKRRIRLVLLVSGVSAASGADNSANDSHSSCPILASAS
AKKEENKRSKSEKGEADYFINSATGFPNLPPEALEESARSLQNOITSSASYLE
GVMSPYKRTISLVSQGDLSIDYFTIPWYDNGSTLEFSQISAKRKEDRTIGNG
LGVQNYGNMLLGNNAFYDYDFRGRRLIGTBAWMDYLKSGNYHPIIDMSQSD
EDPEERBAGMDIRMESWLPEYPOLGAKIYGEQYADYVALEFETDMLQDPHATVG
LETPYPLVYGVGDYKAGTSDNSPYNATNYNOIGPIAQLQDPREVKIIOHSLMSGR
IDFVRNNETITLERKQDPLDVTILMLADATNEHPECVIEDTPEAAYGLECKRTVA
LINHYKTIISAWQAKNNAARTLVMPVKNALALTEGNNNSWNLVLPAMVADTEQRT
ALNTWKVMTLEDEKGNKONGVVEITVODRKTELIVNLTADTDRSDHSEASALAD
GEDVYDLITLSDGSDTRNGNEIYVODRKTELIVNLTADTDRSDHSEASALAD
IASRDKAGATVLTSLTPTGTRMKAKADAGDSNYVYTFGDMLSALNYIYVKA
NPVLTICKEDKHPPVNNAYRTIELMRDKNGVROMSROLTEFEALYDOMEFTGQST
NGHTGALANTNEDLVLPYTNKEAAOKFAANVEDGVGCGIRVYYSOK"
/compelment(9037..9042)
/gene="slnH"
/feature="putative RBS for slnH; RegulonDB:STMS1H002624"
/compelment(9382..9615)
/gene="yfgJ"
/feature="synonym: STM2518"
/compelment(9382..9603)
/gene="yfgJ"
/feature="similar to E. coli orf, hypothetical protein
(AAC75563.1); Blastp hit to AAC75563.1 (83 aa), 63%
identity in aa 13 - 83"
/codon_start=1
/transl_table=11
/product="putative cytoplasmic protein"
/protein_id="AA121412.1"
/db_xref="GI:16421062"
/translation="MEITPCYCHALERNQDPTAHCENAKDFSLQALCPDRCPLQVL
KACGAVDFEONGHGLSKRRVNFVISQ"
/compelment(9610..9615)
/gene="yfgJ"
/feature="putative RBS for yfgJ; RegulonDB:STMS1H002625"
/compelment(9692..11176)
/gene="engA"
/feature="synonym: STM2519"
/compelment(9692..11164)
/gene="engA"
/feature="similar to E. coli putative GTP-binding factor
(AAC75564.1); Blastp hit to AAC75564.1 (503 aa), 97%
identity in aa 14 - 503"
/codon_start=1
/transl_table=11
/product="putative GTP-binding protein"
/protein_id="AA121413.1"
/db_xref="GI:16421063"
/translation="WVPVALVGRPNNGKSTLENRLRTRDALVADEPGLTRDRRYGR
AEVGRRETCIDTGDIDGDEGVEFRMAEQLIAIEADVLFVWDARAGLMPADEAI
AKHLRSRKPFLVANKTDGDPQAVVDYSLGIELEYPIAASGRVLSLEHVL
PMDVDVAQOEEDDAEYMAOEAEONGEEDDEDFDQSPIKAIYGRNVKSTL
TNRLIGERRVYVDMPCGTIRDSIYIPMRDRREYVLDITAVRRKGLTIDVKEFSVI
KTIDAITEDANVLYLTARREISODLSLFLILNLSGSLVYVNNKMGLSQEKQEV
KETIDFLRGTDFARVHTISALHSGVGNLEESYREALDSTRYSTMLRLIMTAV
EDHOPPLVGRVRKLKYAHAGGNPPIVHIGNOVKDLDPDYKRYLAMYFKSLEVMG
TPRIQKEGGENPYANKRNTLTPQMRKKRIMKHIKSK"
/compelment(11171..11176)
/gene="engA"
/feature="putative RBS for engA; RegulonDB:STMS1H002626"
/compelment(11283..12471)
/gene="yfgL"
/feature="synonym: STM2520"

```

## Alignment Scores:

```

Pred. No.: 5,36-132 Length: 23647
Score: 1840.00 Matches: 367
Percent Similarity: 99.46% Conservative: 3
Best Local Similarity: 98.66% Mismatches: 2
Query Match: 98.61% Indels: 0
DB: 1 Gaps: 0

```

US-09-921-992-78 (1-372) x AE008814 (1-23647)

```

OY 1 MethisAnGlnAlaProIleGlnArgArgSerThrArgIleTyrValGlyAsnVal 20
DB 15609 ATGATATACCAAGGCTCCGATTCAACGTAAGAAATCGACAGTATTTCGTTGGAAATGTG 15550
OY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
DB 15549 CCGATTGGCATTGGTCCCGTATGCGCGTCCAGTCGATGACTAACAGCGGTACAGTAC 15490
OY 41 ValGlnAlaThrValAlaSerGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
DB 15489 GTGGAACGACGAGTAAATCAACATCAAGCGCTGAGCGCGTGGCGAGATATTGTCGT 15430
OY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
DB 15429 GTTCTTGACCGAATAAGAGCGCCCGCAACGCTTCAAGCTTATCAACAGCAGATTAC 15370
OY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
DB 15369 GTCCCGCGTGGTGGCATTATTCACCTTCGACTACCGCATTTGGCGTGAAGTACGGAATAC 15310
OY 101 GlyValAspCysLeuAlaArgIleAsnProGlyAsnIleGlyAsnGluGluArgMet 120
DB 15309 GCGCTTGACTGCTTGGCATACACCGCGCAATTCGTTACGAAAGACGATTTCGTATG 15250
OY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAlaAsnAlaGlySer 140
DB 15249 GTGCTGACACTGCGCGCGCAAAAACATCCCTTTCGTATCCGCGTAAACGCGGATCG 15190
OY 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuGluSer 160
DB 15189 CTGGAAGAAAGATCGCAGAGAAATACGCTAGCGCGCAGCGCGGTGCTGGAATCA 15130
OY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
DB 15129 GCCATGCGCCATGTCGATCATCTGATCTCAACTTGTATCATGTTCAAGTACAGCGTA 15070
OY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
DB 15069 AAAGCGTCGATGATTCCTTCCGCTGATCTTATCGTTGCTGCTTGAACATCATCAT 15010
OY 201 GlnProLeuHisLeuGlyIleThrGlnAlaGlyGlyAlaAspSerGlyAlaValLysSer 220
DB 15009 CAGCGCGTCATCTGGGATACCGCAAGCGCGCGCGCTAGCGGCGCTCAAGTCG 14950
OY 221 AlaIleGlyLeuGlyLeuLeuLeuSerGluIleGlyAspThrLeuArgValSerLeu 240
DB 14949 GCGATCGGTTAGCTCTGCTGTATCGGAAGTATCGCGCACACGCTCGGCTATCGTTG 14890
OY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
DB 14889 GCGGCGCAGCCGCGGAGAGATCAAACTCGGTTTCGATATTCGAACTCCTCGCATTT 14830
OY 261 ArgSerArgGlyIleAsnPheIleAlaLysProThrCysSerArgGlnGluPheAspVal 280
DB 14829 CGCGGCGCGGATCAATTTATTCGCTCCGCGACCTTTCTCCGACAGATTGCACTT 14770
OY 281 IleGlyThrValAsnAlaLeuGluGluArgLeuGluAspIleIleThrProMetAspVal 300
DB 14769 ATCGGCAAGGGAACGCGCTGAGCAGCTCGAAGATATCAACACCCGATGACGCTC 14710
OY 301 SerIleIleGlyCysValAlaAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
DB 14709 TCGATCATCGTGTGTAGTAAACGCGCGGCGGAGGCTGTGTATTCACCTGGGCGTGG 14650
OY 321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValAlaGlyAspArgLeuAsp 340
DB 14649 ACGGCGGCAATTAACAAAGCGCGCTGTATTAAGCGCGCTGCTTAAGACAGATCGAT 14590
OY 341 AsnAsnAspMetLysAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
DB 14589 AAGCAGCATATGATCGCGCAGCTTGAATCCGTTATTCGCGGAAAGCAAGTCACTGAT 14530

```





misc\_feature  
/note="Pfam match to entry PF00583 Acetyltransferase (GNAT) family, score 32.50, E-value 9.7e-06"  
complement(3790..3813)  
/gene="STRY2723"  
/note="PS00017 ATP/GMP-binding site motif A (P-loop)"  
complement(4420..5283)  
/gene="STRY2724"  
complement(4420..5283)  
/note="STRY2724"  
/note="Orthologue of E. coli P76563; Fasta hit to P76563 (287 aa), 92% identity in 286 aa overlap. Contains hydrophobic, possible membrane-spanning region"  
/codon\_start=1  
/transl\_table=1  
/product="putative membrane protein"  
/protein\_id="CAD02686.1"  
/db\_xref="GI:16503703"  
/db\_xref="SPTRMBL:082480"  
/translation="MRNQGRESNNVEDRKRPRGSPSLGSGPRIPRKGKGIILLYVYVAGYGVDTLGLTGDPVSQOSTRISPDDEAKFTSYLTATTEDTWOLFQKMGKGYQPKLVMTRGMTGCGAGSGVMPFYCPADGYIYIDISFYDMKNKLGAGDRAQGYIAEYGHVHVKRLGIEPKVRQLQOMASQTEVNRSLQADCFAGVHWSMQOQGLEAGDLEALNAQAIGDRLQOQGGVYVPSFTHTGTSQRYSWFKRFGDSGPAOCNTGKKNF"

RBS  
complement(5287..5292)  
/note="possible RBS"  
complement(5415..6128)  
/gene="STRY2725"

gene  
/note="synonym: purC"  
complement(5415..6128)  
/gene="STRY2725"  
/note="Orthologue of E. coli purC (PUR7\_ECOLI); Fasta hit to PUR7\_ECOLI (237 aa), 95% identity in 237 aa overlap"  
/codon\_start=1  
/transl\_table=1  
/product="phosphoribosylaminoimidazole-succinocarboxamide synthase"  
/protein\_id="CAD02687.1"  
/db\_xref="GI:16503704"  
/db\_xref="SPTRMBL:08XFF69"  
/translation="MKOAEIYRGAKYVSNENDLIVLEPRNDTSAGDARIQDFDKGVNKNFNRIMTKIAGIPIDMERILSTDELYKLEKVPYECYRRRAASISYKRIAGEEMENLPFLDELKNDALHDEPNVSSYCTGFWVQSEMLAKMKELTYKANDVLKLFDDAGLIVDFKLEFGIYKGEVVLGDEFSPDGRIMDKETLIDKMDRFOQLGGLIAYEAVAHRLGVKLD"

misc\_feature  
complement(5421..6122)  
/gene="STRY2725"

misc\_feature  
/note="Pfam match to entry PF01259 SAICAR synt, SAICAR synthetase, score 414.20, E-value 1.2e-120"  
complement(5586..5612)  
/gene="STRY2725"

misc\_feature  
/note="PS01058 SAICAR synthetase signature 2"  
complement(5829..5873)  
/gene="STRY2725"

RBS  
/note="PS01057 SAICAR synthetase signature 1"  
complement(6137..6141)  
/note="possible RBS"  
complement(6303..7337)  
/gene="STRY2726"

gene  
/gene="STRY2726"

Alignment Scores:  
Pred. No.: 2,9e-130 Length: 145050  
Score: 1831.00 Matches: 365  
Percent Similarity: 99.19% Conservative: 4  
Best Local Similarity: 98.12% Mismatches: 3  
Query Match: 98.12% Indels: 0  
Gaps: 0

US-09-921-992-78 (1-372) x AL627275 (1-145050)

QY 1 MethSASGAlaAlaProIleGlnArgGlySerThrArgIleTyrValGlyAsnVal 20  
|||||

Db 64624 ATGCATACACGAGCTCCGATTAACGTAGAGAAATCGACAGCATATTACGTGGAAATGTC 64565  
QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrSerThrArgTrpTrpAsp 40  
Db 64564 CCGATTGGCGATGGCGCCCTTATCGCGCTCAGTCGATCAACACGCTTACACTGAC 64505  
QY 41 ValGlnAlaThrValAsnGlnIleTyrAlaLeuGlnArgValGlyAlaAspIleValArg 60  
Db 64504 GTGACACGACGGTATATCATCATCAAGCCGCTGGAGCGCTTGAGCGAGATATGTGCT 64445  
QY 61 ValSerValProThrMetAspAlaAlaGlnAlaPheTyrSerIleTyrGlnGlnValAsn 80  
Db 64444 GTTCTCTACCGACCAACGAGACGCGCGCAAGCGCTTACAGCTTATCAACAGCGAGTTAGC 64385  
QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuTyrValAlaGluTyr 100  
Db 64384 GTCCCGCTGTTGCCGATATTCCTTGACTACCGCATTCGCGTGAAGAAAGTAGGGAATAC 64325  
QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGlnArgIleArgMet 120  
Db 64324 GCGCTTGACTGCTTGGCGATTAACCCGCGCATATCGTAAACGAACGATTCGCGATG 64265  
QY 121 ValValAspCysAlaArgAspIleAsnIleProIleArgIleGlyValAsnAlaGlySer 140  
Db 64264 GTGGTGAAGCTGCGCGGACAAACATCCCTATCCGCTATCGCGCTAATTCGCGATG 64205  
QY 141 LeuGlnIlyAspLeuGlnIlyAsnIlyGlyIlyGluProThrProGlnAlaLeuLeuGluSer 160  
Db 64204 CTGGAAGAAAGATCTGACGAGAAATACGAGACCGCGCGCGCTGCTCGTGAATCA 64145  
QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheTyrValSerVal 180  
Db 64144 GCCATGCGCCATGTCGATCATCTGATCTCACTTGTATCATGTTCAAAAGTACGCGTA 64085  
QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaValGlnIleAsp 200  
Db 64084 AAAGGCTCGCATGATCTCTCGCTGATCTATCGTTGTTGGCTAAACAGATCATGAT 64025  
QY 201 GlnProLeuHisLeuGlyIleThrGlnAlaGlyGlyAlaAspSerGlyAlaValIlySer 220  
Db 64024 CAGCCGCTGCAATCTGGGATACCGCAACGGCGCGCGCTAGCGGGCGCTTAAGTCG 63965  
QY 221 AlaIleGlyLeuGlyLeuLeuLeuSerGlnIlyIleGlyAspThrLeuArgValSerLeu 240  
Db 63964 GCGATCGGTTTAGCTCTGCTGTATCGGAAGCATCGCGCATCGCTGATCGTTG 63905  
QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260  
Db 63904 GCGCGCGACCGGCTGAGAGATCAAAAGTCGTTGATATCTGMACTGACCTGCGTGT 63845  
QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnIlyPheAspVal 280  
Db 63844 CGCGCGCGCGCATTAATTTATCTGCTGCGCGACTGTTCTTCGTCAGAGATTCGACTT 63785  
QY 281 IleGlyThrValAsnAlaLeuGlnArgLeuGlnArgLeuAspIleIleThrProMetAspVal 300  
Db 63784 ATCGGACAGGTGACAGCGCTGAGCAGCGTGAAGATATCATCAACCCGATGAGCTGC 63725  
QY 301 SerIleIleGlyCysValValAsnIlyProGlyGlnAlaLeuValSerThrLeuGlyAla 320  
Db 63724 TCGATCATCGCTTGTGTGTGAACGGCGCGCGGCGAGCGGTGTATCAACGCTGGCGCTG 63665  
QY 321 ThrGlyIlyAsnIlyLysSerGlyLeuTyrGlnAspGlyValArgIlyAspArgLeuAsp 340  
Db 63664 ACGGGCGGCAATTAAGAAAGCGGCTGTATGAAGCGCGTCCGTAAGACGACTCAT 63605  
QY 341 AsnAsnAspMetIleAspIleLeuGlnAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360  
Db 63604 AACGACGATATGATCGCGAGCTTGAATCCGCTATTCGCGCGAAAGTAACTCACTGAT 63545  
QY 361 GluAlaArgArgIleAspValGlnGlnValGluLys 372  
Db 63544 GAAGCGCGCTGGATGACGCTGTCAGGCTTGAAGAA 63509

```

RESULT 15
LOCUS       AE016835               300431 bp    DNA          linear    BCT 21-MAR-2003
DEFINITION  Salmonella enterica subsp. enterica serovar Typh1 Ty2, section 2 of
ACCESSION   AE016835
VERSION     AE014613
KEYWORDS
SOURCE      Salmonella enterica subsp. enterica serovar Typh1 Ty2
ORGANISM    Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1 (bases 1 to 300431)
AUTHORS     Deng, W., Liou, S.-R., Plunkett III, G., Mayhew, G.F., Rose, D.J.,
            Burland, V., Kodoyianni, V., Schwartz, D.C. and Blattner, F.R.,
            Comparative Genomics of Salmonella enterica Serovar Typh1 Strains
            Ty2 and CT18
JOURNAL     J. Bacteriol. 185 (7), 2330-2337 (2003)
MEDLINE     22531367
PUBMED      12644504
REFERENCE   2 (bases 1 to 300431)
AUTHORS     Deng, W., Liou, S.-R., Plunkett, G. III, Mayhew, G.F., Rose, D.J.,
            Burland, V., Kodoyianni, V., Schwartz, D.C. and Blattner, F.R.,
            Direct Submission
            Submitted (25-SEP-2002) Laboratory of Genetics, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
            location/Qualifiers
            1. 300431
            /organism="Salmonella enterica subsp. enterica serovar
            Typh1 Ty2"
            /mol_type="genomic DNA"
            /strain="Ty2"
            /serovar="Typh1"
            /sub_species="enterica"
            /db_xref="taxon:209261"
            complement(67..486)
            /gene="yfiG"
            /locus_tag="t0261"
            complement(67..486)
            /note="corresponds to STY2842 from Accession AL513382:
            Salmonella typh1 CT18"
            /codon_start=1
            /transl_table=11
            /product="thioredoxin 2"
            /protein_id="AA067987.1"
            /db_xref="GI:29136421"
            /translation="MNTVCTHCQAINRIRGDLKQDAKCGRCGHELFDEGVINATGET
            LKRLKDDLPVVIDENAPWCGPCRNPAIFEDVAERSGKVRFKVYNTAEARELSARF
            GRSIPMIMIFRHHGVDMNGAVRPAPDSWLNEL"
            690..1727
            /gene="yfiF"
            /locus_tag="t0262"
            690..1727
            /gene="yfiF"
            /locus_tag="t0262"
            /gene="yfiF"
            /locus_tag="t0262"
            /note="corresponds to STY2841 from Accession AL513382:
            Salmonella typh1 CT18"
            /codon_start=1
            /transl_table=11
            /product="putative RNA methyltransferase"
            /protein_id="AA067988.1"
            /db_xref="GI:29136422"
            /translation="MNDELKNSGKGVKVVYRSDDSDKRTNHPRTGKGGPAPKASRT
            DGRPARBERNNOSDRKHKTSPPRTVSRAPGDETPKVDHGSGSEFDEPLAR
            ORAETRYVGENACOLFQSRPDAIVRAWFIOGVTPRPERALRMWAAHRAVHYVDEA
            ELAKSGTEHHGGCVFLIKRNGTIVKQWVCOAODCVYLAIEDVANHNKAGMMRSC
            AHFGKVVYVODAAALLESAAIRTAEGGAENHOPITGSGSYDVLDLDFKQAGTVVTS
            SGRGALSTTIPKAMVYLGRHYDLEPAAREPDLCKINGTGNVESLNVSAIVG
            LIAEMWROKKA
            complement(1843..2532)

```

```

CDS
            /gene="ung"
            /locus_tag="t0263"
            complement(1843..2532)
            /gene="ung"
            /locus_tag="t0263"
            /note="corresponds to STY2840 from Accession AL513382:
            Salmonella typh1 CT18"
            /codon_start=1
            /transl_table=11
            /product="uracil-DNA glycosylase"
            /protein_id="AA067989.1"
            /db_xref="GI:29136423"
            /translation="MATELIMHVLADENQOQPYFINTLTAVAGERSGITYVPPQDV
            ENFRTEIGDVAVILGDPPYHGPQAGLAFLSRPGIAPPSLVNMYKLEASIPG
            EYRPAHGYLESARQGVLLNTLVLRACQASHASIGMETFDKVIYSLINRGCV
            FLIMGSHACKGAIIDPQRHILKAPHPSPLSAHRQFFCGNHFALNMLEQNGERTI
            DWTPVLPASE"
            2851..3234
            /gene="yfiD"
            /locus_tag="t0264"
            2851..3234
            /gene="yfiD"
            /locus_tag="t0264"
            /note="corresponds to STY2839 from Accession AL513382:
            Salmonella typh1 CT18"
            /codon_start=1
            /transl_table=11
            /product="conserved hypothetical protein"
            /protein_id="AA067990.1"
            /db_xref="GI:29136424"
            /translation="MTGQITQIKRANDDILNSFWLDSKRGARCTVAKSGRAEDRY
            AVSKGEIYRREIPHEKREYVREGQHLNVLRRELTLEDVAKHEKYPQLTIKVS
            YAVRNSLTPEQORDIARTFEST"
            complement(3296..3883)
            /gene="yfiK"
            /locus_tag="t0265"
            complement(3296..3883)
            /gene="yfiK"
            /locus_tag="t0265"
            /note="corresponds to STY2838 from Accession AL513382:
            Salmonella typh1 CT18"
            /codon_start=1
            /transl_table=11
            /product="putative membrane protein"
            /protein_id="AA067991.1"
            /db_xref="GI:29136425"
            /translation="WTPMLLSFWTYTITLTPGPNNILASNTANGFROSIRVLA
            GMSIGFLVMLCAGIAFSLAVIDPAIHLISWGAAYILMAKIASPDADEKVRP
            KFGVFWSEGLQFVNVKIIIGTALSTFVLDPQDALMWVIGVSILLALIGFGVWC
            ALAGHLFORAFHYHQRLNIILALLVCVARIY"
            3986..4868
            /gene="yfiE"
            /locus_tag="t0266"
            /note="corresponds to STY2837 from Accession AL513382:
            Salmonella typh1 CT18"
            /pseudo
            complement(4901..6235)
            /gene="srmB"
            /locus_tag="t0267"
            complement(4901..6235)
            /gene="srmB"
            /locus_tag="t0267"
            /note="corresponds to STY2836 from Accession AL513382:
            Salmonella typh1 CT18"
            /codon_start=1
            /transl_table=11
            /product="ATP-dependent RNA helicase SrmB"
            /protein_id="AA067992.1"
            /db_xref="GI:29136426"
            /translation="MTVTFSELESDSLDALODKGFTRPTAIOAAIPALDGRV
            LGSAPITGKTAVALPALOHLLDPRKKSPPRIILITPRRELAAOVADARLALAH
            THLDATITGGVAYAMNHAERESENQDIYVATIGLLOITKEENFDCRAVEFTLIIDED
            RMLDMGFRQDIEHAGETIRNRKQTLMSATLEGAITQFAERLLEDPEVASNPSTRE

```

RKIIHWTYRADNPEHKKVALLKHLKODDTRTSIVYRKREHVELAETRLAGINC  
 YLGEAAQIKRNEGIRKLTGDRVNAVIVATVAARGIDIPVSHVINDFRSGOTIYH  
 RIGRTGAGRRGTALISLVEAHDLILKTRITIEELKAVIDELRTTAPSEKLTG  
 KPSKRYLAKRAEKKKEKKEKPRVKRYRDTKNIGKRRKPSGTRKMOSSSE"  
 6365..7102  
 /gene="yfic"  
 /locus\_tag="t0268"  
 6365..7102  
 /gene="yfic"  
 /locus\_tag="t0268"  
 /note="corresponds to STY2835 from Accession AL513382;  
 Salmonella typhi CT18"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AA067993.1"  
 /db\_xref="GI:29136427"  
 /translation="MSQSGSVLRNNGFTFKQFVYAHDRCAKMGVTDGILGAMPAVD  
 VKRIIDIGTGGGLALMAORFSDSVPADELDAAGAMQAEVNAHSPMHRITVHT  
 DDIORMAPROTVREDDIISNPYEPVEGSTPQREORATATDHOTLLAIADCLT  
 EDGFCVLPBQIENAFQCALNMGWHRLRTDVAENEARLPHVILAFSPQAECS  
 DRLVIRGSDNVTSTALTOAFILFM"  
 complement(7087..8709)  
 /gene="nab"  
 /locus\_tag="t0269"  
 complement(7087..8709)  
 /gene="nab"  
 /locus\_tag="t0269"  
 /note="quinolinate synthetase B; corresponds to STY2834  
 from Accession AL513382; Salmonella typhi CT18"  
 /codon\_start=1  
 /transl\_table=11  
 /product="L-aspartate oxidase"  
 /protein\_id="AA067994.1"  
 /db\_xref="GI:29136428"  
 /translation="MMTPPELSCDYLITGSGAAGISLARLEAKHKVIVLSKGPVSES  
 STFYAOGGIAAVFEDTDSIAHVEDTLIAGAGICORHAEVVASNAKTCVOMLDQV  
 LEVDIVQNGKESYHLIREGSHRIRLHADATGEVETLVSAQNHPIQYIERS  
 NAVDLISDKIGLPGPRVGVAMINRWKEVETCHASVYLATGSAKVQYITNP  
 ISSGDGIMAMRACGRVNLNENOPHPTALYHQPANFLTEALGEGAYLKRPDGS  
 EMPVDERGELAPRDIVARAIDHEKRRGACOMFLDISHKPDPEYRHPMTIYALID  
 LGMDLTKERPIVPAARTCGSVYVDDGRVDYGLVAGISVYGLHAGNRMASNSL  
 LECIVYGSAAAMDIDRMPVSHSVVLEFAMDESREKADDERVIOHNMHELRLMNDY  
 VGIVTRRLERALARITMLQOEIDEVYANFRSVNNLLELNLVVOAELIVCAMMRK  
 ESRGHFTLDYFQOLAESPSILSPLTHINK"  
 9134..9709  
 /gene="rpoE"  
 /locus\_tag="t0270"  
 9134..9709  
 /gene="rpoE"  
 /locus\_tag="t0270"  
 /note="sigma-24; corresponds to STY2833 from Accession  
 AL513382; Salmonella typhi CT18"  
 /codon\_start=1  
 /transl\_table=11  
 /product="RNA polymerase sigma-E factor"  
 /protein\_id="AA067995.1"

## Alignment Scores:

Pred. No.: 7,58e-130 300431  
 Score: 1831.00 Matches: 365  
 Percent Similarity: 99.19% Conservative: 4  
 Best Local Similarity: 98.12% Mismatches: 3  
 Query Match: 98.12% Indels: 0  
 DB: 1 Gaps: 0

US-09-921-992-78 (1-372) x AE016835 (1-300431)

QY 1 Methiasasnglnalaproilleginaraglysserthrarglletyvalglyasnval 20  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 83116 ATGCATACACAGGCTCCGATTCACGTAGAAAATCACAACGATATTACGTGGGAATGTG 83175  
 QY 21 Proileglyaspelvalaproillealvalglnsermethrasnthrarghthrthrasp 40

||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 83176 CCATATGGCATGGCGCCCTATCGCGCTCAGTCATGACTAACCGGCTACCGACTGAC 83235  
 QY 41 Valgluaalathrvalasnglnilleysalaleugluarvalglylaaspllevalay 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 83236 GTGGAAAGCGACGGTTATATCATGATCAAGCCCTGGAGCGGCTTGCGAGATATTGTCGT 83295  
 QY 61 Valservalprothrmelaspalaalaglualaphelysleuileysglnvalasn 80  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 83296 GTTTCCTACCGACCATGAGAGCGCGCGAGCGCTCAAGCTTATCAACAGCAGGTTACG 83355  
 QY 81 Valproleuvalalaaspllehisphisasptyrargllealaleuysvalalaglutyr 100  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 83356 GTCCCGCTGGTGGCGATATTCACCTTCGACTACCGCATTCGCTGAAGAGTAGCGAATAC 83415  
 QY 101 Glyvalaspcysleuarglleasnproglyasnilleysnglnluarvalleargmet 120  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 83416 GCGCTTACCTGCTGCTGCTATTAACCCCGCAATATCGGTACAGAAAGCAAGCATTCGATG 83475  
 QY 121 Valvalaspcysalalaraspllyasnilleproilleargllevalasnalaglyser 140  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 83476 GTGGTGACATGCGCGCGCGCAAAAACATCCCTATCCGTATCGCGCTTAATGCGCGATCG 83535  
 QY 141 Leuuluysasplleuglnulysstyrglyulprothrproglualaleuugluser 160  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 83536 CTGGAAGAAAGATCTGCAGAAATATACGAGAGCCGACGCGCGCGCTGCTGATCA 83595  
 QY 161 Alamelarghisvalasphisleuasparglleuasnphesapginphelysvalserval 180  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 83596 GCCATGGCGCATGTCGATCATCTGCATGCTCAACTTGTATGCTTCAAGAGTAGAGCTA 83655  
 QY 181 Lysalaserasparvalphelualalvalglnserlyrargleuualalysglnleasp 200  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 83656 AAACGCTCCGATGATTCCTGCTGCGTGAAGCTTATGCTTGTGCTTAACAGATCGAT 83715  
 QY 201 Glnproleuhsleuglyllethrglualaglylalaraserglyalavalysser 220  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 83716 CAGCCGCTGCATCTCGGGATATACCGAAGCGGCGCGGCTACCGGGCGGTAAATCG 83775  
 QY 221 Alaileglyleuglyleuileuileusergluylilleysapthreulargvalserleu 240  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 83776 GCGATCGGTTTAGGCTGCTGCTGTATCGGAAGCATCGCGACGCTGCGGTATCGTGG 83835  
 QY 241 Alaalaaspprovalglnulilleysvalaglypheasplleleuysserleuarglle 260  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 83836 GCGCGCGACCGGAGGAGATCAATCGGTTTCGATTCGTAAGTCGCTCGCAT 83895  
 QY 261 Argserargglylleasnphellealacysprothrlysserargglnulphespyal 280  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 83896 CCGCGCGCGCGCATATTCATCGCTGCGCGACCTGCTTCGCAAGAGTTGACGCTT 83955  
 QY 281 Ileglythrvalasnaleuglnulargleuulnaaspllelethrprometaspyal 300  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 83956 ATCGGACGCGGACCGCTGAGCAGCGCTGGAAGATATCATACCCCATGACGCTC 84015  
 QY 301 Serilleileglycysvalvalasnglyproglylualaleuvalserthrleuglyval 320  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 84016 TCGATCATCGGTGTGTGTGTGTGACAGCGCGCGCGCAAGCGCTGATTCACGCTGGCGTG 84075  
 QY 321 Thrglyglyasnlylserserlyleuylarglyvalarglyaspargleuasp 340  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 84076 ACGGCGCGCAATTAAGAAAGCGCGCTGATGAAGACGCGCGCTTAAGACAGACTCGAT 84135  
 QY 341 Asnasnaspmetilleasplleuugluarvalleargllealalysalaserglnleuasp 360  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 84136 AACGACGATATGATCGCGAGCTTGAATCCGCTATTCGCGGAGAAAGTAACTGAT 84195

Search completed: August 15, 2003, 00:56:40

Fri Aug 15 14:34:48 2003

Job time : 5342 secs

us-09-921-992-78.rge

\_\_\_\_\_

Run on: August 14, 2003, 21:49:35 ; Search time 356 Seconds

(without alignments)  
2820.761 Million cell updates/sec

Title: US-09-921-992-78

Sequence: 1 MHNQAPIQRKSTRITVGNV.....RAKASQLDEARRIDVQVEK 372

### Scoring table:

Xgapop	10.0	Xgapext	0.5
Ygapop	10.0	Ygapext	0.5
Fgapop	6.0	Fgapext	7.0
Delop	6.0	Delext	7.0

Searched: 2552756 seqs, 1349719017 residues

total number of hits satisfying chosen parameters: 5105512

Maximum DB seq length: 200000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Command line parameters:

```

0-FCG12.1/USFPO.spool1/US09921992/r/unat.12082003.094910.17333/app_query.fasta.1.51918
MODEL=1/rname+ .p2n.model -DEV=xlh
-DB-N.GeneSeq.19Jun03 -OEM=fastab -SUFFIX=mg -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -STRAT=1 -END=1 -MATRIX=blissum62 -TRANS=human40.cd1
-LIST=45 -DOCLIGN=200 -THR_SCORE=Pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUT=TFP.plo -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09921992.@CG1.1.1.312.g/unat.12082003.094910.17333 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -OSBLOCK=100 -LONLOC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELDP=6 -DELETE=7

```

Database : N\_Geneseq\_19Jun03:\*

1:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed

and is derived by analysis of the total score distribution

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1866	100.0	1119	21	AAA95478	E. coli essential
2	1866	100.0	1119	21	AAAH8705	E. coli rRN essent
3	1866	100.0	1119	22	AAAC82653	E. coli gcpE DNA.
4	1866	100.0	1119	24	AAAD31203	Escherichia coli g
5	1603.5	85.9	1830121	17	AAK42063	Haemophilus influ
6	1282	62.7	64081	22	ABAF2857	Buohnera sp. genom
7	1166	68.5	33140	24	ABOC67966	Genomic fragment #
8	868	46.5	1107	24	ABOC69933	Listeria monocyrog
9	868	46.5	2944528	24	ABAO03041	Listeria monocyrog
10	868	46.2	6157	24	ABO70939	Listeria monocyrog
11	861.5	46.2	6157	24	ABK74401	Bacillus monocyrog
12	832	44.1	1083	24	ABO81842	Bacillus licheniflo
13	814.5	43.6	349980	24	ABO68334	Blifidobacterium 1o
14	812	43.5	994	24	AAAX14017	Listeria monocyrog
15	794	42.6	1180	23	AAAS59573	H. pylori GHPD 76
16	785.5	42.1	35829	19	AAAH67170	Propionibacterium
17	761	40.8	1134	22	AAAH67170	C. glutamicum codin
18	761	40.8	1179	25	AAAC00649	C. glutamicum deri
19	761	40.8	349980	22	AAAH68350	C. glutamicum codin
20	761	40.3	349980	22	AAAH68351	C. glutamicum codin
21	751.5	40.3	1161	25	ABK71126	Mycobacterium tube
22	751.5	40.3	1164	22	AAAH20566	Mycobacterium tube
23	751.5	40.3	4403765	22	AAAP9683	Mycobacterium tube
24	751.5	40.3	4411529	22	AAAP9683	Mycobacterium tube
25	682	36.5	5484	23	AAAT30062	DNA encoding novel
26	534	28.6	1263	25	ABK239756	N. gonorrhoeae nuc
27	527	28.2	60873	21	AAAB1469	N. meningitidis pa
28	527	28.2	349980	21	AAE71610	Neisseria meningit
29	527	28.2	1437668	21	AAAB1469	N. meningitidis B
30	510	27.3	1857	23	AAAS0066	DNA encoding novel
31	490	26.3	1038602	20	AAAZ01425	Complete genome se
32	470.5	25.2	273254	21	AAAC81914	Chlamydia pneumoni
33	470.5	25.2	1230025	20	AAAX91990	Nucleotide sequenc
34	462	24.8	2520	24	AAAB31201	Arabidopsis thalia
35	460	24.7	2109	22	AAAC82654	P. falciparum gcpE
36	446	23.9	3147	21	AAAC82996	P. falciparum gcpE
37	428.5	23.0	507	24	AAAT75580	Human ORF135
38	428.5	23.0	507	24	ABN766194	Human ORF143 cDNA
39	383	20.5	435	24	AAAB66784	Helicobacter pylor
40	339	18.2	670	24	AAAD31221	Zea mays partial g
41	326	17.5	596	24	AAAD31222	Zea mays partial g
42	321	17.2	584	24	AAAD31220	Zea mays partial g
43	300	16.1	4467	10	AAAN92428	Sequence encoding
44	292	15.6	33675	24	AAAD31202	Oryza sativa gcpE
45	277.5	14.9	2535	20	AAAZ0674	Polynucleotide seq

## ALIGNMENTS

```

RESULT 1
AAA95478
ID      AAA95478 standard; DNA; 1119 BP.
XY

```

AC AAA95478

DT 27-FEB-2001 (first entry)  
YY

DE E. coli essential gene *gcpe*  
XX

**KM**

H. influenza; S. pneumoniae; H. pylori; B. burgdorferi; neisseria;

M. tuberculosis; antibiotic; ds.

OS Escherichia coli

XX MO200061793-A2.  
 XX  
 XX 19-OCT-2000.  
 XX  
 XX 07-APR-2000; 2000MO-EP03135.  
 XX  
 XX 09-APR-1999; 99EP-0107031.  
 PR 04-FEB-2000; 2000EP-0102131.  
 XX  
 XX (GPCB-) GPC BIOTECH AG.  
 XX  
 XX Loferer H, Jacobi A;  
 PI  
 XX WPI; 2000-687048/67.  
 DR  
 XX Identifying antibacterial compounds, comprises identifying an  
 PT antagonist or inhibitor of the expression of a gene encoding a  
 PT polypeptide essential for bacterial growth or survival -  
 XX  
 PS Claim 1; Fig 1; 75pp; English.  
 XX  
 CC The present invention relates to antagonists and inhibitors of 24  
 CC bacterial genes and proteins. The proteins are thought to be essential  
 CC for growth in several species of bacteria (including *S. pneumoniae*, *B.*  
 CC *burgdorferi*, *H. influenza* and *H. pylori*). The proteins and coding  
 CC sequences shown in the specification can be used to identify antagonists  
 CC and inhibitors which can be used in disease treatment and pesticides. In  
 CC particular, they can be used against *M. tuberculosis*. The present  
 CC sequence is one of the genes of the invention.  
 XX  
 SQ Sequence 1119 BP; 272 A; 209 C; 317 G; 251 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 7.12e-104 Length: 1119  
 Score: 1866.00 Matches: 372  
 Percent Similarity: 100.003 Conservative: 0  
 Best Local Similarity: 100.003 Mismatches: 0  
 Query Match: 100.003 Indels: 0  
 DB: Gaps: 0

US-09-921-992-78 (1-372) x AAAS9478 (1-1119)

OY 1 MethisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20  
 DB 1 ATGCATTAACCAACGCTCGATTCATCAAGTAAATCAACACGATTTAGCTGGGAAATG 60  
 OY 21 ProIleGlyAspGlyAspProIleAlaValGlnSerMetThrAspThrArgThrAsp 40  
 DB 61 CCGATTGGCGGATGGTGGCCATCCGCTACAGTCATGACCAATACGCGTACGACAGC 120  
 OY 41 ValGlnAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60  
 DB 121 GTGCAAGCAACGCTCATCAATCAAGCGCTGGAACGGCTGGCGATATCGTCGT 180  
 OY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnIleValAsn 80  
 DB 181 GATCCGCTACCAACGCTGACGCGGAGAGGCTTCACAACTCATCAACAGCAGGTTAAC 240  
 OY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100  
 DB 241 GTGCCCGCTGGTGGTGAATCCACTTCGACTGCGATTCGCGTGAAGTACGGAATAC 300  
 OY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluIleArgIleArgMet 120  
 DB 301 GGGCTGATTCCTGCGTATTAACTGGAATATGCTAAATGACAGCGTATTCGATG 360  
 OY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140  
 DB 361 GTGGTGAATCTGGCGGATTAACATTCGATCGATTCGATTCGATTCGATTCGATTCG 420  
 OY 141 LeuGluAspLeuGluGluLysTyrGluProThrProGlnAlaLeuGluGluSer 160

DB 421 CTGAAAAAAGATCTGCAGAAAGAAAGTATGCGAACCAGCCGACGCGTGTGCGAATCT 480  
 OY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180  
 DB 481 GCCATGCGGTCAATGATTCATCTCATGCTCCGCTGAACCTTGATGATCAAACTCAGCGTG 540  
 OY 181 LysAlaSerAspValPheLeuAlaValGlnSerTyrArgIleLeuAlaLysGlnIleAsp 200  
 DB 541 AAGCGCTGACGCTCTCTCCGCTGTGAGTGTTCGTTTCGTTGCTGCGCAAAACAGATCAT 600  
 OY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyIleAlaArgSerGlyAlaValLysSer 220  
 DB 601 CACCGCTGCATCTCGGATCAACCGGAGCGGTGTGCGGAGCGGCGGAGTAAATCC 660  
 OY 221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240  
 DB 661 GCCATTGTTAGTCTGCTGCTGTGTAAGCATGCGGACAGCGTGGCGATTCGCTG 720  
 OY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260  
 DB 721 GCGCGCGATCCGCTCGAAGAGATCAAAAGTCGCTTCGATATTTGAAATCGCTGCTATC 780  
 OY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnIlePheAspVal 280  
 DB 781 CGTTGCGAGGATCAATCTCATGCTGCTCCGACCTGTTCCGCTCAGGATTTGATGTT 840  
 OY 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300  
 DB 841 ATGGTACGGTTAACCGCGCTGAGCAACGCTCGAAGATATCATCTCGATGAGACGTT 900  
 OY 301 SerIleIleGlyCysValValAsnGlyProGlyGlnAlaLeuValSerThrLeuGlyVal 320  
 DB 901 TCGATTTCGCGCTGCGTGTGAATGGCCAGGTGAGCGCTGTTCTTACACTCGCGCTG 960  
 OY 321 ThrGlyLysAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340  
 DB 961 ACCGCGGCAACAAAGAAAGCGGCTGTATGAAGATGCGTGGCAAAAGCCGTCGTGAC 1020  
 OY 341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360  
 DB 1021 AACCAACATATGATGACCAACGCTGGAAGCAACGCAATTCGTCGGAAGGCAATCGCTGAC 1080  
 OY 361 GlnAlaArgArgIleAspValGlnGlnValGluLys 372  
 DB 1081 GAAGCGCTCAATTGACGTTCAACAGCTTGAAAAA 1116

RESULT 2  
 AAA88705  
 ID AAA88705 standard; DNA; 1119 BP.  
 XX  
 AC AAA88705;  
 XX  
 DT 05-FEB-2001 (first entry)  
 XX  
 DE E. coli FUN essential gene gcpE.  
 XX  
 KW FUN gene; gcpE gene; essential gene; antibacterial; antibiotic;  
 KW screening; infection; therapy; antagonist; surrogate marker; ds.  
 XX  
 OS Escherichia coli.  
 XX  
 PD EP1043403-A1.  
 XX  
 PD 11-OCT-2000.  
 XX  
 XX 09-APR-1999; 99EP-0107031.  
 PF  
 XX 09-APR-1999; 99EP-0107031.  
 PR  
 XX (GPCB-) GPC GENOME PHARM CORP AG.  
 PA  
 XX Not given;  
 PI  
 XX



DR WPI: 2000-640125/62.

XX Identifying antagonists of the expression of gene encoding bacterial  
PT growth polypeptide useful for treating bacterial infections or  
PT diseases, by evaluating transcription of the gene in the presence of  
PT test molecule -

XX Claim 1; Page 25; 55pp; English.

XX The present sequence is that of the Escherichia coli FUN gene  
XX gcpE, which encodes a protein that is essential for bacterial  
XX growth or survival. gcpE is 1 of 22 E. coli genes (see  
XX AA88692-713) identified as being essential (there is no deletion  
XX genotype). These 22 genes fulfill criteria for being attractive  
XX antibacterial targets: hypothetical open reading frames coding for  
XX essential functions (mutation is lethal for growth in rich media);  
XX broad conservation (orthologues are present in a wide range of  
XX bacteria, including Haemophilus influenzae, Streptococcus pneumoniae,  
XX Helicobacter pylori and Borrelia burgdorferi); and low toxicity  
XX potential in higher organisms (mostly no orthologues were identified  
XX in Saccharomyces cerevisiae). An antagonist or inhibitor of the  
XX expression of an essential gene or of its function provides the key  
XX for antibacterial therapy. The invention provides methods for  
XX identifying such antagonists or inhibitors. These involve  
XX contacting a bacterial cell comprising an essential gene with a  
XX candidate antagonist or inhibitor, and testing whether contact leads  
XX to cell growth inhibition or/and cell death. The method allows the  
XX development of new broad spectrum antibiotics. A conditional mutant  
XX of an essential gene can be used to induce a lethal phenotype in  
XX bacteria for the analysis of surrogate markers.

SQ Sequence 1119 BP; 272 A; 279 C; 317 G; 251 T; 0 other;

Alignment Scores:

Pred. No.:	7.12e-184	Length:	1119
Score:	1866.00	Matches:	372
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	21	Gaps:	0

US-09-921-992-78 (1-372) x AAA8705 (1-1119)

QY 1 MethAsnGlnAlaProIleGlnArgGlySerThrArgIleTyrValGlyAsnVal 20  
DB 1 ATGCATACCAAGGCTCCAAATCAACGTAACAAATCAACGTAATTCATCGTGGAGATGTG 60  
QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAspThrArgThrAsp 40  
DB 61 CCGATTGGGAGTGTCTCCATCGCCGATGACATGCAATACGCGTACGACAGAC 120  
QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60  
DB 121 GTCGAAGCAACGCTCAATCAATCAAGCGCTGGAAGCGGTGGCCGATATCGCGCT 180  
QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnIleValAsn 80  
DB 181 GATTCGTCACCAAGATGAGCGCGAGAGGCTTCAAACTCATCAAAACAGCAGGTAAAC 240  
QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100  
DB 241 GTCCGCGTGTGGTGGACATCCACTTCGACATTCGCGATTCGCTGAAGTACGGAAATAC 300  
QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120  
DB 301 GCGCGTCAATGTCCTCGATTAACCTGCAATATCGGAATGAGAGCGATTCGCAATG 360  
QY 121 ValValAspCysAlaArgAspLysAsnIleProIleGlyIleGlyValAlaAsnAlaLysSer 140  
DB 361 GTGGTGTACTGTGCGCGGATTAACATCCGATTCGATTCGCGTAAACCGCGGATTCG 420  
QY 141 LeuGluLysAspLeuGlnIleLysTyrGlyLysIleProIleAlaLeuLeuGluLysSer 160

DB 421 CTGGAAGAAAGATCTGCAGAAAGATATGCGCAACCGACGCCGCGCTTGTGTGATCT 480  
QY 161 AlMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180  
DB 481 GCCATGGCTATGTATGATCATCTCGATCGGCTGACCTTCATCATGTTCAAAAGTCAGCGTG 540  
QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuAlaLysGlnIleAsp 200  
DB 541 AAAGCGTCACTCTTCCCTCGCTGTGATGCTTATATGCTTGTGCGCAAAACAGATCGAT 600  
QY 201 GlnProLeuHisLeuGlyIleIleHisGluAlaGlyValAlaArgSerGlyAlaValLysSer 220  
DB 601 CACCGCTGATCTGGGATCCAGAGATCAAGAGCGGTGCGCGCAGCGGCGAGTAAATCC 660  
QY 221 AlaIleGlyLeuGlyLeuLeuLeuSerGlyIleGlyAspPheLeuArgValSerLeu 240  
DB 661 GCCATGCTATGATGCTGTGCTGTGCTGGAAGGATCGGCGACACGCTGGCGATCGCTG 720  
QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260  
DB 721 GCGGCGGATCCGCTCGAAGATCAAGATCGGTTTGATATTTGAATTCGCTGCTATAC 780  
QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnLysPheAspVal 280  
DB 781 CGTTCCGAGAGGATCAACTTCATCGCTGCGGACCTGTTCCGCTCAGCAATTTGATGTT 840  
QY 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300  
DB 841 ATCGTACGCTTAACCGCTGCGTGAAGCCTGGAAGATATCATCTCCGATGAGCGTT 900  
QY 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320  
DB 901 TCGATATATCGCTGCTGCTGTGAATGCGCCAGGTGAAGCGCTGTCTTACACTCGCGCTC 960  
QY 321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValAlaArgLysAspArgLeuAsp 340  
DB 961 ACCGCGGCGCAACAAGAGCGGCTCTATGAAGATGGCGTGGCAAAACCGCTCGGAC 1020  
QY 341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerLeuAsp 360  
DB 1021 AACACGATATGATGTCACCAAGCTGGAAGCAGCATTCGCGCAAAAGCAGTACGCTGAC 1080  
QY 361 GluAlaArgArgIleAspValGlnGluValGlyLys 372  
DB 1081 GAAGCGCGTCAATGACGTTACAGCAGGTTGAAAAA 1116

RESULT 3  
AAC82653 standard; DNA; 1119 BP.  
AAC82653;  
15-MAR-2001 (first entry)  
E. coli gcpE DNA.  
Isoprenoid; gcpE; yf8b; antimicrobial; transgenic plant; agriculture;  
antimycotic; antiparasitic; antiviral; fungicidal; herbicidal; ds.  
Escherichia coli.  
WO200072022-A1.  
30-NOV-2000.  
20-MAY-2000; 2000WO-EP04592.  
21-MAY-1999; 99DE-1023567.  
21-MAY-1999; 99DE-1023568.  
(JOMA/) JOMAA H.  
Jomaa H;



XX (MONS ) MONSANTO TECHNOLOGY LLC.  
 PA Boronate A, Campos N, Rodriguez-concepcion M, Rohner M, Seeman M;  
 PI Valentin HE, Venkatesh TV, Venkatramesh M;  
 XX WPI: 2002-227151/28.  
 DR P-PSDB: AAE19653.  
 XX  
 PT gcpe nucleic acid which is an essential gene of the methyl-D-erythritol  
 PT phosphate pathway, encoding a fully defined GCPE protein which is  
 XX useful for increasing levels of tocopherol substrates in plants  
 XX  
 XX Claim 3; Page 117-119; 155pp; English.  
 CC The invention relates to gcpe nucleic acid molecule, an essential gene  
 CC of methyl-D-erythritol phosphate (MEP) pathway) that encodes rice,  
 CC Arabidopsis thaliana or Escherichia coli GCPE protein. gcpe is useful  
 CC for producing a transgenic plant such as Brassica campestris, B. napus,  
 CC canola, castor bean, coconut, cotton, crame, linseed, maize, mustard,  
 CC oil palm, peanut, rapeseed, rice, safflower, sesame, soybean, sunflower,  
 CC or wheat with an increased isoprenoid (tocopherol) compound level. The  
 CC expression of GCPE protein in organisms increases the level of  
 CC tocopherol substrate such as isopentenyl diphosphate and dimethylallyl  
 CC diphosphate biosynthesis. Transgenic organisms overexpressing GCPE  
 CC protein can nutritionally enhance food and feed sources. Overexpression  
 CC of GCPE protein in transgenic plant may provide tolerance to stresses  
 CC e.g., oxidative stress tolerance such as to oxygen or ozone, UV  
 CC tolerance, etc. gcpe may be used to obtain nucleic acid molecules from  
 CC the same species, and to obtain nucleic acid homologues. gcpe is also  
 CC used as or primers. The recombinant vectors are used in plant  
 CC transformation or transfection. gcpe also act as markers capable of  
 CC detecting polymorphisms such as single nucleotide polymorphisms (SNPs).  
 CC gcpe is also used to determine the level or pattern of expression of  
 CC the protein. The present sequence is Escherichia coli gcpe gene.  
 CC  
 XX  
 SO Sequence 1119 BP; 272 A; 279 C; 317 G; 251 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 7, 12e-184 Length: 1119  
 Score: 1866.00 Matches: 372  
 Percent Similarity: 100.00% Conservatave: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 24 Gaps: 0  
 US-09-921-992-78 (1-372) x AAD31203 (1-1119)  
 QY 1 Methlaaanglnalaproilleglnargarglyssertharqglyerthvalglsnval 20  
 DB 1 ATGCATACCAAGGCTCCAAATCAAGTAAATCAACGATTTTAACTGGGAATGTG 60  
 QY 21 Proilleglaaspclvalaprollealvalaaglnsermetthasntharqthrhrasp 40  
 DB 61 CCGATGGGGAGTGGTCCATCGCCGTACAGTCCATACCAATACCGGTACGACAGAC 120  
 QY 41 VALGLUALFTHVALASNGINILEYLSALAEUGLUARGVALGILVALASPILEVALARG 60  
 DB 121 GTGGAAGCAACGGTCAATCAATCAAGGCGTGGAGACGGTGCATATCGCCGT 180  
 QY 61 VALSERVALPROTHMETASPAIALAAGLUALAPHELVALLEILYSGINLVALASN 80  
 DB 181 GATCCGTCACCAACATGACGCGCAGACAGAGCTTCAACTCAACACAGAGGTAAAC 240  
 QY 81 VALPROLEUVALAASPILLEHISPHASPTYRARGILEALALEULYSVALAAGLUTYR 100  
 DB 241 GTGCGCGTGGTGGTGCATCCACTTCGACATTCGATTCGCTGAAAGTACGGGAATAC 300  
 QY 101 GLYVALASPCYSLEUARGILEASNPARGLYASNILEGLYASNGILGLUARGILEARGMET 120  
 DB 301 GCGCGTCGATGTCTCGATTAACCCCTGGCAATATCGGTAAAGAACGCGATTCGCATG 360  
 QY 121 VALVALASPCYSALARGASPLYSANILEPROILLEARGILEGLYVALASALAEGLYSER 140

DB 361 GTGCTTACCTGTGGCGCATTAACATTCGATCCGATTTGCGCTTAACGCCGATG 420  
 QY 141 LeuglulysaspheuglnglulysTYRGLYGLYProthrProglinalaleuengluser 160  
 DB 421 CTGGAAGAAAGATGTGCAAGAAAGTATGGGAACCGACGCCGACGGCTTGTGGATCT 480  
 QY 161 AlameTarghlshValasphlsleuaspargleuasnpheaspclnphelaysValSerVal 180  
 DB 481 GCCATCGCATCTTATCATCTCATCTCGATCCGCTGAACCTGCATCATGATCAAGCACGCTG 540  
 QY 181 LysAlaSeraspValpHeuAlaValgluserTYRArgleuAlaLysglnleasp 200  
 DB 541 AAGCGTTCGACCTCTTCCTCGCTGTGAGTCTTATCGTTTGTGGCAAAACAGATCGAT 600  
 QY 201 GlnProleuHlsleuglylIethrGluAlaGlyglYAlaArgserGlyAlaVallysSer 220  
 DB 601 CACCGCTGCATCTGGGGATCACCGAAGCGGTGGCGCGACGGGCAATMAATCC 660  
 QY 221 AlaIlleGlyLeuGlyLeuLeuLeuSerGlyGlyIleGlyAspThrLeuArgValSerLeu 240  
 DB 661 GCCATTGGTTCAGTCTGCTGCTGCTGCTGAAGCATCGCGACACGCTGCGCGATCGCTG 720  
 QY 241 AlaAlaAspProValglulglulileysValglYpHeaspIlleuLeuLysSerLeuArgIle 260  
 DB 721 GCGCGCGATCCGCTGCAAGATCAAGTGGTTCGATATTTGAATCGCTGCTATC 780  
 QY 261 ArgSerArgGlylLeasnpHealIealacyProthrCysSerArgGlnlupHeaspVal 280  
 DB 781 CTTTCGCGAGGGATCAACTTCAATCGCTCCGACCTGCTCCGCTGAGAAATTTGATGTT 840  
 QY 281 IleGlyThrValAsnAlaLeuGlulnArgleuGluAspIlleThrPrometaspVal 300  
 DB 841 ATCGGACGCTTACGCGCTGAGCAACCCCGGAAGATATCATCTCCGATGACGCT 900  
 QY 301 SerIleIleGlyCysValValasnglyProglYgluAlaLeuValSerThrleuGlyVal 320  
 DB 901 TCGATTATCGGCTGGGTGTGATGGCCAGGTGAGGCCCTGTTTCTACACTCGGCGTTC 960  
 QY 321 ThrnglyGlyAsnLysSerGlyLeuTYRgluAspGlyValArgLysAspArgleuasp 340  
 DB 961 ACCGGCGGCAACGAAAGACCGCCTCATGAAAGATGGCGTCCGCAAGAACCGCTCGAC 1020  
 QY 341 AsnAsnAspMetIleAspGlnleuGluAlaArgIleArgAlaLysSerGlnleuasp 360  
 DB 1021 AACACGATATGATGACACGCTGAGACACCGCATTCGTGCAAAAGCCAGTCAGCTGAC 1080  
 QY 361 GlualaArgArgIleaspValGlnlValGluLys 372  
 DB 1081 GAAGCGCGTCGAATGACGTTACAGCTTGAAGAAA 1116  
 RESULT 5  
 AAT42063  
 ID AAT42063 standard; DNA; 1830121 BP.  
 XX AAT42063;  
 AC  
 XX  
 DT 14-SEP-1999 (first entry)  
 XX  
 DE Haemophilus influenzae complete genome sequence.  
 XX  
 KW Genome; bacterium; Haemophilus influenzae; computer readable medium;  
 KW expression modulating fragment; regulation; gene expression; vector;  
 KW organism; open reading frame; ORF; ds.  
 XX  
 OS Haemophilus influenzae.  
 XX  
 FN W09633276-A1.  
 XX  
 PD 24-OCT-1996.  
 XX  
 PF 22-APR-1996; 96WO-US05320.

PR 07-JUN-1995; 95US-0487420  
 PR 21-APR-1995; 95US-0426790  
 PR 07-JUN-1995; 95US-0476102  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (UYJO) UNIV JOHN HOPKINS  
 XX  
 PI Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;  
 DR WPI; 1996-485782/48.  
 XX  
 PT Haemophilus influenzae Rd genome recorded on computer readable  
 PT medium - useful for identifying commercially important nucleic acid  
 PT fragments by homology searching  
 XX  
 PS Claim 1; Page 77.2-77.1090; 1291pp; English.

CC This sequence represents the complete genome sequence of the bacterium  
 CC Haemophilus influenzae strain Rd. The invention relates to a computer  
 CC readable medium (CRM) having recorded upon it the complete H. influenzae  
 CC nucleotide sequence (1), a representative fragment of (1) or a nucleotide  
 CC sequence at least 998 identical to (1). By providing the full-length  
 CC genomic sequence in a computer readable form, it is possible to identify  
 CC commercially important nucleic acid fragments and expression modulating  
 CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to  
 CC regulate the expression of a nucleic acid molecule. Vectors and altered  
 CC organisms comprising the predicted ORFs can be used to produce any of the  
 CC polypeptide fragments of the H. influenzae Rd genome.  
 XX

Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other:

Alignment Scores:  
 Pred. No.: 2,6e-152 Length: 1830121  
 Score: 1603.50 Matches: 320  
 Percent Similarity: 93.663 Conservative: 20  
 Best Local Similarity: 88.158 Mismatches: 22  
 Query Match: 85.938 Indels: 1  
 DB: 17 Gaps: 1

US-09-921-992-78 (1-372) x AATG063 (1-1830121)

QY 4 GlnAlaProIleGlnArgLysSerThrArgIleTyrValGlyAsnValProIleGly 23  
 DB 390972 CAGCCAACTATTACGCTGCAATGACAAATAATTGTGTGGAAATGTACCAATTGGT 391031  
 QY 24 AspGlyAlaProIleGlnValGlnSerMetThrAsnThrArgThrThrAspValGluAla 43  
 DB 391032 GGGGATGCCCTTATGGCTGCATCATGACAAATFACCGCACACATGATGGAGCG 391091  
 QY 44 ThrValAsnGlnIleGlyAlaLeuGluArgValGlyAlaAspIleValArgValSerVal 63  
 DB 391092 ACAGTTGCTCAATTAATTAATCATTAAGACGTGTGGCAGATATTGCTGCTATCTGTT 391151  
 QY 64 ProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGluValAsnValProLeu 83  
 DB 391152 CCACAAATGAGTCGCGACACCATTTAAACAATTAACAACAAGTGAATGCTTCGGCTC 391211  
 QY 84 ValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAsp 103  
 DB 391212 GTAGCAGATATTCATTCGACATCTGATCGGCTTAAAGTGGCAGAAATATGAGAGTAT 391271  
 QY 104 CysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValAlaAsp 123  
 DB 391272 TGTTCAGTATCAATGCTGCAACATTTGCTGCTGAGAGTGGCTGCTGCTGTGAT 391331  
 QY 124 CysAlaArgAspLysAsnIleProIleArgIleGlyValAlaAlaGlySerLeuGluLys 143  
 DB 391332 TGTGCGGAGACAAATATTCGATTCGATTTGGTGTAATGACAGGCTTTAGAAAAA 391391  
 QY 144 AspLeuGlnGluLysArgIleGluProThrProGlnAlaLeuLeuGluSerAlaMetArg 163  
 DB 391392 GATTTCAGAAAAAATATGCGCAGACCAAGCCGCTTGTGAATCCGATTCGCT 391451

QY 164 HisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSer 183  
 DB 391452 CATGTAGAAATTTCTAGATGCTCTTAATCTGCATGACATTAAATGAGCGTAAAAAGCCTCC 391511  
 QY 184 AspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeu 203  
 DB 391512 GATGTATCTTACGGGTGATCTTATCCGTTACTGCTTAACCAATTAACAGCCTTTA 391571  
 QY 204 HisLeuGlyIleThrGlnAlaGlyValAlaArgSerGlyValAlaLysSerAlaIleGly 223  
 DB 391572 CATTTAGCCATTACAGAGAGAGGCGGACCGGCTGCTGCATTAATATCGAGTGGGT 391631  
 QY 224 LeuGlyLeuLeuLeuSerGlnGlyIleGlyAspThrLeuArgValSerLeuAlaIleAsp 243  
 DB 391632 TTAGCAATGTTATTAAGCTGAGGCGATGCGGATACACTACGCGCTCTTTGGCGCAGAT 391691  
 QY 244 ProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArg 263  
 DB 391692 CCTGTAGAGAAACCAAGTCGCTTGTGATTTGAAATCTTTACGATTCGTTCAAGA 391751  
 QY 264 GlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThr 283  
 DB 391752 GGATTTAACTTATTATGCTTGCCCAACCTGTTCCGCAAGAAATTTGATGTAATCGGTACA 391811  
 QY 284 ValAsnAlaLeuGlnGluArgLeuGluAspIleIleThrProMetAspValSerIleIle 303  
 DB 391812 GTAAATGGCTAGAACCAACGCTTGAAGATATTTACCAACCAAGATGATCTATTATC 391871  
 QY 304 GlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGly 323  
 DB 391872 GGTTCGTAGTGAATGCTGCTGCGGAGGCACTGCTCCGATCTCGGCGTAAACGGCGGT 391931  
 QY 324 AsnLysLysSerGlyLeuTyrGluAspGlyValArg---LysAspArgLeuAspAsn 342  
 DB 391932 AACAAAAAAGCGGTTATTATCTTACGAGAGAACCAAAAAAGCGTTTGTATACGAA 391991  
 QY 343 AspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAla 362  
 DB 391992 GATATAGTGAACCAATTTGAAGCAAAAATTTGCTGCAAAAGTCGACCAACAAAGATCCAAA 392051  
 QY 363 ArgArgIle 365  
 DB 392052 AACAGAAAT 392060  
 DB  
 RESULT 6  
 ABA92787 standard; DNA; 640681 BP.  
 ID ABA92787  
 XX  
 AC ABA92787;  
 XX  
 DT 27-MAR-2002 (first entry)  
 XX  
 DE Buchnera sp. genomic DNA SEQ ID NO:1.  
 XX  
 KW Buchnera; cockroach-symbiotic bacterium; cockroach extermination;  
 XX circular; ds.  
 OS Buchnera sp.  
 XX  
 PN JP2001292771-A.  
 XX  
 PD 23-OCT-2001.  
 XX  
 PF 07-APR-2000; 2000JP-0107160.  
 XX  
 PR 07-APR-2000; 2000JP-0107160.  
 XX  
 PA (RIKA) RIKAGAKU KENKYUSHO.  
 XX  
 DR WPI; 2002-126043/17.  
 XX  
 PT A genomic DNA of cockroach-symbiotic bacterium

200 ASPG1NPROLEUHTSLKGLYLIEHFGIUAAGLYGLYLAARGSERGLYALAVALLYS 21

The present invention relates to a "Memory-Independent" Claim 1, Page 191-199; 545pp; English.

comprising of a combination of 41 nucleic acid molecules (see the present invention relates to a moraxella catarrhalis genomic library comprising of a combination of 41 nucleic acid molecules (see





SO Sequence 1107 BP; 348 A; 178 C; 255 G; 326 T; 0 other;

# Alignment Scores:

Pred. No.:	2,34e-80	Length:	1107
Score:	868.00	Matches:	178
Percent Similarity:	67.68%	Conservative:	67
Best Local Similarity:	49.17%	Mismatches:	115
Query Match:	46.52%	Indels:	2
DB:	24	Gaps:	2

US-09-921-992-78 (1-372) x AB067966 (1-1107)

```

QY 10 ArgLysSerThrArg---ileYrValGlyAsnValProIleGlyAspGlyAlaProIle 28
DB 19 CCGGAAAACACTGCGCCAGTCCAAAGGGAATTATTAATTTAGTATGCTGAGTGAAGATTA 78
QY 29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGlnAlaThrValAsnGlnIle 48
DB 79 ACTATCCAAAGCATGACTACTACAAAGACATGATGATGCGAACACAGACAGCAAAAT 138
QY 49 LysAlaLeuGlnArgValGlyAlaAspIleValAlaArgValSerValProThrMetAspAla 68
DB 139 CACCGATTAGAAAGACCTGCTTCAGATTGTCGAGTTCCTGCTCTGATGACGATGCA 198
QY 69 AlaGlnAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHis 88
DB 199 GCGAATGCTCTTAGTCCATCAAGAAAAAGATTCATATTCCTGCTGCGCATATTCAT 258
QY 89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
DB 259 TTGGATTACCGACTACACTTAAAGCTATTGATGCGAGTGTGACAAATTCGATTAAT 318
QY 109 ProGlyAsnIleGlyAsnGlnArgMetValAlaAspCysAlaArgAspLys 128
DB 319 CCGGTACATGTCGTGCGCGTGCAGTGAAGAAAGGTATGCTGCTGAAGCAAAA 378
QY 129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGlnLysAspLeuGlnLys 148
DB 379 AATATTCCAATCCGATTTGGGTTAATGCTGATTTGAAAGAAATTTATTCAAAA 438
QY 149 TyrGlyGluProThrProGlnAlaLeuLysGlnSerAlaMetArgHisValAspHisLeu 168
DB 439 TATGTTACCTACTGCTGACGGAATGAGAAAGCGACTTGCCCATTTAAATTC 498
QY 169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
DB 499 GAAGATTGTGATTTATATATCATTTCTTTGAAAGCTTCATGATGATTAAGCA 558
QY 189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
DB 559 ATTGAAGCTTATGATTAAGCTTATGCTGCGCATTTATTCCTGATCTCGGAATTACA 618
QY 209 GlnAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeu 228
DB 619 GAATCTGGACACAAATTTCTGAGGAATAAAAGTCGCGCTTTAGGACCGATATCC 678
QY 229 SerGlnGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGlnLysIle 248
DB 679 AATTGGGCAATGGAAATACATTAGAGATCTTTGAGTGCATGATCCGTCGGAAGAAATA 738
QY 249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
DB 739 AAGAGTGGCCCGGGAATTTAAATCAATTTGCTCTTCCTCAAGATCCCGCAATGCTTATC 798
QY 269 AlaCysProThrCysSerArgGlnGlnPheAspValIleGlyTyrValAsnAlaLeuGln 288
DB 799 TCCTGGCTACTTGGGCGTGAATAGATGATTTAATTCGATGCGTATATAAGTGAA 858
QY 289 GlnArgLeuGlnAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn 308
DB 859 AATTACATAGCAAAAGATTGCAATTCGATTAAGTAGCCGTCCTGGCTGCGGTCAAC 918
QY 309 GlyProGlyGlnAlaLeuValSerThrLeuGlyValIleThrGlyAsnLysSerGly 328

```

```

DB 919 GGCCTGGAGAGACCTGCGCAAGCCGATATCGGAATTCGTTCAAACCGAGAGCCCTT 978
QY 329 LeuTyrGlnAspGlyValArgLysAspArgLeuAspAsnAsnMetIleAspGlnLeu 348
DB 979 CTTTTAGACATGCTAAATTTATTCGAAAGATCCGGAAGCTTTATGATAGCAACTT 1038
QY 349 GlnAlaArgIleArgAlaLysAlaSerGlnLeuAspGlnAlaArgArgIleAspValGln 368
DB 1039 AAGAAAGAAATTCATATTTTGGCAGAAAGATTT---TTTGTGAAGAAATATGATTTGAA 1095
QY 369 GlnVal 370
DB 1096 AGCCTT 1101

```

```

RESULT 9
AB069993
ID AB069993 standard; DNA; 1107 BP.
XX
AC AB069993;
XX
DT 29-AUG-2002 (first entry)
XX
DE Listeria monocytogenes EGDe DNA sequence #205.
XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
XX infection; ds.
XX Listeria monocytogenes EGDe.
XX
OS WO200228891-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-FR03061.
XX
PR 04-OCT-2000; 2000FR-0012697.
XX
PA (INSP ) INST PASTEUR.
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI Kunst F, Glaser P;
XX
DR WPI; 2002-332479/37.
XX
PT New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators -
XX
PS Claim 16; SEQ ID 2806; 180pp; French.
XX
CC The present invention relates to nucleic acid sequences
CC (AB067188-AB071212) from Listeria sp. The sequences are useful as probes
CC and primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in
CC anti-Listeria vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pcr_sequences.
XX
SO Sequence 1107 BP; 348 A; 178 C; 255 G; 326 T; 0 other;

```

## Alignment Scores:

Pred. No.:	2,34e-80	Length:	1107
Score:	868.00	Matches:	178
Percent Similarity:	67.68%	Conservative:	67
Best Local Similarity:	49.17%	Mismatches:	115
Query Match:	46.52%	Indels:	2
DB:	24	Gaps:	2

us-09-921-992-78 (1-372) x AB069993 (1-1107)

```

OY  ArglyserThrArg---IleTyValGlyValAsnValProIleGlyAspGlyAlaProIle 28
Db  CGGAAAACACTCCGCCAGTCAGTGGTAATTAATGCTATGCTGAGTGAAGAAATTA 78
OY  29 AlaValGlnSerMetThrAsnThrArgThrAspValGluAlaIleThrValAsnGlnIle 48
Db  79 ACATACCAAGACATGACTACTACCAAGACATGATCGCAAGCAAGCAAGTGCAGAAAT 138
OY  49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
Db  139 CACCGATTAGAGAAAGCTGGTTGTCAGATGTGGAGTGGTCTTCTCGATGAACGTCGA 198
OY  69 AlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHis 88
Db  199 GCGAATGCTCTTAGCGCATCAAGAAAGATTCATATTCGCTGTCGACAGATATTCAT 258
OY  89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
Db  259 TTGATATACCGACATGACCTTAAGCTATTGATCAGAGTGTGGCAAAATTCGGATTAAT 318
OY  109 ProGlyAsnIleGlyAsnGluGluArgIleArgMetValAlaAspCysAlaArgAspLys 128
Db  319 CCTGGAACATTTGGTCCGCCGTCGATCGGGTGAAGAAAGTGTAAATGCTGTAAGCAAAA 378
OY  129 AsnIleProIleArgIleGlyValAlaAsnAlaGlySerLeuGluLysAspLeuGlnLys 148
Db  379 AATATTCATCCATCCGATATGGGGTTAATGCTGTATGTAAGAAAGAAATTAATTCAAAA 438
OY  149 TyrGlyGluProThrProGlnAlaLeuLeuGlnLysSerAlaMetAlaGlnHisValAspHisLeu 168
Db  439 TATGTTTACCTTACCTACGTCGAGAGATGATGACAAAGTGCACCTTGCATTAATAATTC 498
OY  169 AsparGluLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
Db  499 GAACTATTAGATTATTTATGATATCATCTTCTTGAAGGCTTCGATGATGAATTTACCA 558
OY  189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
Db  559 ATTGAAGCTTATGATTAAGCTAAGTGCATTTAATTCCTCTCTCATCTCGGAATTTACA 618
OY  209 GluAlaGlyGlyAlaArgSerGlyValAlaValLysSerAlaIleGlyLeuGlyLeuLeu 228
Db  619 GAATCTGGTACCAATTTGCTGGAGAGATTAAGTGCCTGCTTAAGCAAGCATCTACTC 678
OY  229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGlnGlnIle 248
Db  679 AGTTTGGCATTTGGAATATACATATACAGATCTTTGAGTCTGATCTGTGGAAGAAATA 738
OY  249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
Db  739 AAAGTGGCCCGGAGAGTTTAAATCAATTTGCTTCTTCCTCAATGCCGCAATGCTTATC 798
OY  269 AlaCysProThrCysSerArgGlnLysPheAspValIleGlyThrValAsnAlaLeuGlu 288
Db  799 TCCTCCCTACTTCTGGCTGCTGATAGATTAATTAATTCGATGCTATCAATTAAGTGA 858
OY  289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn 308
Db  859 AATTTCATATGCAAAAGTTGAAAGTTCGATTAAGTAAAGTACCGCTGCTGCTCGGTCAAC 918
OY  309 GlyProGlyGluAlaLeuValSerThrLeuGlyValIleThrGlyValLysLysSerGly 328
Db  919 GGCCTGGAGAGAGCTCGGCAAGCCGATATCGCAATTTGCTGTAACGAGAAAGCCCTT 978
OY  329 LeuTyrGluAspGlyValArgLysAspArgLeuAspAsnAspMetIleAspGlnLeu 348
Db  979 CTTTTTAGACATGATGAATAATTAATTCGAAAGTACCGGAGAGCTATTATGTTACGAACT 1038
OY  349 GluAlaArgIleAlaArgAlaLysAlaSerGlnLeuAspGluAlaArgArgIleAspValGln 368

```

```

Db  1039 AAGAAAGAAATGATATTTTGGCAGAGAAATTT---TTTGTGAAGAAATATGATTGGAA 1095
OY  369 GlnVal 370
Db  1096 AGCCTT 1101
RESULT 10
ABAB03041
ID  ABAB03041 standard; DNA: 2944528 BP.
AC  ABA03041;
XX  05-FEB-2002 (first entry)
DE  Listeria monocytogenes EGD-e genome sequence.
XX  Listeria monocytogenes EGD-e genome sequence.
XX  Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW  vitamin B12; bacterial infection; disease; ds.
XX  Listeria monocytogenes.
XX  W0200177335-A2.
XX  18-OCT-2001.
XX  11-APR-2001; 2001WO-FR01118.
XX  11-APR-2000; 2000FR-0004629.
XX  (INSP ) INST PASTEUR.
XX  Buchrieser C, Frangeul L, Couve E, Rusniok C, Eshti H, Deloux P,
PI  Dussauget O, Chetoui H, Nedjari H, Glaser P, Kunst F, Cossart P,
PI  Daniels J, Goebel M, Krefit J, Kuhn M, Ng E, Vazquez-Boland JA,
PI  Dominguez-Bernal G, Garrido-Garcia P, Trierrez-Martinez A, Amend A,
PI  Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L,
PI  Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N,
PI  Madueno E, De Pablos B, Wehlund J, Kaerst U, Entian K, Hauf J,
PI  Rose M, Voss H;
XX  WPI, 2002-010914/01.
XX  DR
XX  Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT  and prevention of Listeria and related bacterial infections, and
PT  related polypeptides -
XX  PS
XX  Claim 1; SEQ ID No 1; 192pp; French.
XX  The present sequence is the genome sequence of Listeria monocytogenes
CC  EGD-e. This sequence and fragments of this sequence are useful for
CC  selecting probes and primers for detecting genes in L. monocytogenes and
CC  related organisms, and to study genetic polymorphisms and other genomes.
CC  proteins (ABAB7297-ABAB50149) expressed from the present sequence are
CC  useful for raising specific antibodies, identification of L.
CC  monocytogenes and related organisms, and for biosynthesis and
CC  biodegradation, especially biosynthesis of Vitamin B12. This sequence and
CC  proteins encoded by it are also useful for selecting compounds that
CC  regulate gene expression and cell replication and modulate L.
CC  monocytogenes-related diseases. In addition, this sequence and proteins
CC  encoded by it are useful in pharmaceutical and vaccines compositions for
CC  the treatment or prevention of infections by L. monocytogenes and related
CC  organisms.
CC  Note: The sequence data for this patent did not form part of the printed
CC  specification, but was obtained in electronic format directly from WIP
CC  at ftp.wipo.int/pub/published_pcl_sequences.
XX  SO  Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;

```

Alignment Scores: 9 72e-76 Length: 2944528  
 Pred. No.: 868.00 Matches: 178  
 Score: 67.68% Conservative: 67  
 Percent Similarity: 49.17% Mismatches: 115  
 Best Local Similarity:



```

Query Match: 46.52% Indels: 2
DB: 24 Gaps: 2
US-09-921-992-78 (1-372) x ABA03041 (1-2944528)

QY 10 ArglySerThrArg---lleTyrValGlyAsnValProIleGlyAspGlyAlaProIle 28
DB 1474820 CGCGAAAACACTGGCCGACCAAGTGGTAATTACTGTGTGCTAGTGAAGAAATTA 1474879

QY 29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGlnAlaThrValAsnGlnIle 48
DB 1474880 ACATCCAAAGCATGACTACTACAAAGACACATGATGTCGACAGCAGTACGAGAAAT 1474939

QY 49 LysAlaLeuGluArgValGlyAlaAspIleValAlaGlyValSerValProThrMetAspAla 68
DB 1474940 CACCGATTAGAACAGCTGGTGTGCTGATGTCGAGTTGCTGTCTGTATGACAGTCCA 1474999

QY 69 AlaGluAlaThrLeuLeuLeuLeuGlnGlnValAsnValProLeuValAlaAspIleHis 88
DB 1475000 GCGCAATGCTCTTGTGCTGCAACAAAGAAATTCATTCCTGCTGTGCGAGATATTCAT 1475059

QY 89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
DB 1475060 TTGGATTACCGACTACAGCTAAGCTATGATGACAGTGTGACAAATTCGATTAAAT 1475119

QY 109 ProGlyAsnIleGlyAsnGluGluArgGlyLeuGlyMetValValAspCysAlaArgAspLys 128
DB 1475120 CCTGGTACATGTTGGTCCGCGTATCGGAGAAAGTGGTAAATGCTGCTAAAGCAAAA 1475179

QY 129 AsnIleProIleArgIleGlyValAlaAsnAlaGlySerLeuGlnLysAspLeuGlnLys 148
DB 1475180 AATATTCCAATCCGCTTTTGCGGTTAAATGCTGGTGTAGAAAAGAAATTAATTCAAAAA 1475239

QY 149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
DB 1475240 TATGGTTACCTACTCTGACGGAATGAGAAAGGCACTTGCCCATTTAAATTCCTC 1475299

QY 169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
DB 1475300 GAAGATTATGATTATTTATGATATCATTTCTTTGMAAGCTTCGTGATGTAATTAGCA 1475359

QY 189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuIleThr 208
DB 1475360 ATTGAAGCTTATGATTAACCTAGTCCGCAATTAAATATTCCTGATCTCGGAATTACA 1475419

QY 209 GluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeu 228
DB 1475420 GAATCTGTACACAAATTTGCTGAGAGAAATAAAAAGCGCTGTTAGGACCGATCTC 1475479

QY 229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluLeu 248
DB 1475480 AGTTTGGCATTGGAAATCATATTACGATCTTTGAGAGCGTCAATCTGGAAGAAATA 1475539

QY 249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyLysAsnHelle 268
DB 1475540 AAAAGGGCCCGGGAAGTTTAAATCATTTGCTCTTCTCCAAATCCGCAATGCTTATTC 1475599

QY 269 AlaCysProThrCysSerArgGlnLysPheAspValIleGlyThrValAsnAlaLeuGlu 288
DB 1475600 TCCTCCCTACTTGGCTGCTAGATGATGATTAATTCGATCGCAATGAATGGA 1475659

QY 289 GlnArgLeuGluAspIleIleThrPrometAspValSerIleIleGlyCysValValAsn 308
DB 1475660 AATTACATACGCAAGATTGAATTCCTGATTAAGACCGTGGCTGTCGCGCTCAAC 1475719

QY 309 GlyProGlyGluAlaLeuValSerThrLeuGlyValIleThrGlyLysLysSerGly 328
DB 1475720 GGCCCTGGAGAAAGCTCGGAGCCGATATCGGAATTCGTGTCTCAACGAGAACGCTT 1475779

QY 329 LeuTyrGluAspGlyValAlaArgLysAspArgLeuAspAsnAspMetIleAspGlnLeu 348
DB 1475780 CTTTATTAGACATGTAAATTAATTCGAAAGTACCGAAGCTATATGATGACGAATCTT 1475839

```

```

QY 349 GluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAlaArgArgIleAspValGln 368
DB 1475840 AAGAAAGAAATTTGATTTTGGCAGAAATTT--TTTGTCAAGAAATATGATTTGAA 1475896

QY 369 GlnVal 370
DB 1475897 AGCCTT 1475902

RESULT 11
ABQ70939
ID ABQ70939 standard; DNA; 6157 BP.
XX
AC ABQ70939;
XX
DT 29-AUG-2002 (first entry)
XX
DE Listeria monocytogenes 4b contig DNA sequence #881.
XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
XX infection; ds.
XX
OS Listeria monocytogenes 4b.
XX
PN WO200228891-A2.
XX
PD 11-APR-2002.
XX
PE 04-OCT-2001; 2001WO-FR03061.
XX
PR 04-OCT-2000; 2000FR-0012697.
XX
PA (INSP ) INST PASTEUR.
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI Kunst F, Glaser P;
XX
DR WPI; 2002-332479/37.
XX
PT New genomic sequences from Listeria species, useful for detection,
XX treatment and prevention of infection, also related polypeptides,
XX antibodies and modulators.
XX
PS Claim 14; SEQ ID 3752; 180bp; French.
XX
XX
CC The present invention relates to nucleic acid sequences
CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
CC and primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in
CC anti-Listeria vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6157 BP; 1731 A; 1065 C; 1209 G; 2150 T; 2 other;

Alignment Scores:
Pred. No.: 1.12e-78 Length: 6157
Score: 861.50 Matches: 175
Percent Similarity: 68.66% Conservative: 66
Best Local Similarity: 49.86% Mismatches: 109
Query Match: 46.17% Gaps: 1
DB: 24 Indels: 1

US-09-921-992-78 (1-372) x ABQ70939 (1-6157)

QY 10 ArglySerThrArg---lleTyrValGlyAsnValProIleGlyAspGlyAlaProIle 28
DB 3123 CGCGAAAACACTGGCCGACGATCAAGTATGATTAATTACTGTGTGCTAGTGAAGAAATTA 3182

```

```

OY 29 Alavalginsermettthasnthrargfthrthraspvalglualathraslasnglnlle 48
DB 3183 ACTATCCAAAGCATGAGTACTACAAAGACATGATGTCGAAAGCAACAGTAGCGGAATTT 3242
OY 49 Lysalalaleugluarvdlglialaaspillevaalarvalservalprothmetaspala 68
DB 3243 CACCGATTAGAAAGGAGGGTGTGCACATTTGTGCGAATGCTGTCTGTGATGACCGTCA 3302
OY 69 AlaciuallapheylsleuileylsnglnvalasnaValproleuValalaspillehis 88
DB 3303 GCCCAATCTCTTACTGATCAGAACAAAGATTATCTCTCTGTTGGAGATATTCAT 3362
OY 89 PheasptyrargileuileuileuValalaglufyrilysaspysleuargileasn 108
DB 3363 TTTGATTAACCGACATGACCTTAAGCATATGATGATGATGATGATGATGATGATGAT 3422
OY 109 Progllyasnilleglialynglnluarvleargmetvalvalaspysalalaspys 128
DB 3423 CCTGTAAACATCGCTCGCGTGCATCGGCGGAAAGGTTAATGCTGTAAGCAAA 3482
OY 129 AsnilleprolleargldleuValasnaaglyserleuglulysaspysleu 148
DB 3483 AATATTCATCCATCCGTAAGGGGTTAATGCTGTGTTAGTAAAGAAATATTATCAAAA 3542
OY 149 TyrclygluprothrproglinalaleuileuileuSeralemetarghisvalaspysleu 168
DB 3543 TACGGTTACCTACTGCGGAAAGGAAATGTAAGTGCATGATGATGATGATGATGATGAT 3602
OY 169 AspargleuasnpheaspnglnpheyvalservallysalaaseraspvalphleuVala 188
DB 3603 GAAATATTTAGATTTTATGATATCATGTTCTTGTGAAGGCTTCGATGATGATGATGAT 3662
OY 189 ValgluserTyrargileuileuileuileuileuileuileuileuileuileuileu 208
DB 3663 ATGAGAGCTTATGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3722
OY 209 Glualaglyglialaargserglyalavallysseralalleuileuileuileu 228
DB 3723 GAATCTGCTAGCGAGTGTGCTGAGGATGATGATGATGATGATGATGATGATGATGAT 3782
OY 229 SeruglyylleglyaspthrleuargvalserleuValalaspysleuileuileu 248
DB 3783 AGTTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3842
OY 249 Lysvalilglypheaspilleglyserleuargileargserargglyileasnphelle 268
DB 3843 AAAGTGGCCCGGAGGAGGATTAAGATCTTGGCTTCTGCAATGCCCATGCTTATC 3902
OY 269 AlacysprothrCysSerarglnluarvleuileuileuileuileuileuileu 288
DB 3903 TCTTCCCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3962
OY 289 Gluargleuileuileuileuileuileuileuileuileuileuileuileuileu 308
DB 3963 AATTACATACGACATGATTAAGCTCCGATTAAAGCGCTGCTGCTGCTGCTGCTGCT 4022
OY 309 GlyprogllyglialaleuValserthrleuileuileuileuileuileuileu 328
DB 4023 GGTCCCGGAGAGCTGCGGAAACGATATCGGAATGCTGCTGCTGCTGCTGCTGCT 4082
OY 329 LeuTYRgluaspilValarglysaspargleuaspasnaspmetilleaspilneu 348
DB 4083 CTTTATTAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4142
OY 349 Glualaargileargalalysalaserleu 359
DB 4143 AAGAAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4175

```

RESULT 12  
ABK74401  
ID ABK74401 standard; DNA; 1083 BP.  
XX ABK74401;  
AC ABK74401;

```

XX 13-AUG-2002 (first entry)
DE Bacillus licheniformis genomic sequence tag (GST) #1692.
XX Differential gene expression; genomic sequenced tag; GST;
KW altered culture condition; environmental stress;
KM physiological provocation; ds.
XX Bacillus licheniformis.
OS W0200229113-A2.
PN 11-APR-2002.
PD 05-OCT-2001; 2001WO-US31437.
PE 06-OCT-2000; 2000US-0680598.
PR 27-MAR-2001; 2001US-279526P.
XX (NOVO) NOVOZYMES BIOTECH INC.
PA (NOVO) NOVOZYMES AS.
XX Berka R, Clausen IG;
PI WPI: 2002-416684/44.
XX Monitoring differential expression of several genes in first Bacillus
PT cell relative to expression of same genes in one or more second
PT Bacillus cells, by using substrate containing Bacillus genomic
PT sequenced tag array -
XX Claim 4; SEQ ID NO 1692; 200pp; English.
XX The invention describes a method of monitoring differential expression of
XX genes in a first Bacillus cell relative to expression of the genes in
XX other Bacillus cells, comprising hybridising labelled nucleic acid probes
XX isolated from Bacillus cells to a substrate containing array of Bacillus
XX genomic sequenced tags (GST), examining the array, and determining
XX relative gene expression by an observed hybridisation reporter signal of
XX a spot in the array. The method is useful for measuring the expression of
XX genes in a first Bacillus cell relative to expression of the same genes
XX in one or more second Bacillus cells. The method is useful for monitoring
XX global expression of several genes from a Bacillus cell, discovering new
XX genes, identifying possible functions of unknown open reading frames and
XX monitoring gene copy number variation and stability. Monitoring changes
XX in expression of genes may be used to provide a representation of the way
XX in which Bacillus cells adapt to changes in culture conditions,
XX environmental stress or other physiological provocation. Extensive
XX follow-up characterisation is unnecessary, when one spot on an array
XX equals one gene or one open reading frame, since sequence information is
XX available. This sequence represents a genomic sequence tag (GST) used in
XX the method of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at
XX ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 1083 BP; 314 A; 278 C; 265 G; 226 T; 0 other:

```

Alignment Scores:

Pred. No.:	1.34e-75	Length:	1083
Score:	822.00	Matches:	162
Percent Similarity:	68.45%	Conservative:	55
Best Local Similarity:	51.10%	Mismatches:	100
Query Match:	44.05%	Gaps:	0
DB:	24		

US-09-921-992-78 (1-372) x ABK74401 (1-1083)

```

OY 9 Argarglyyserttharvleuileuileuileuileuileuileuileuileuileu 28
DB 24 CCGTCAAAAGCGCTGCGTAAAGTGGAGGACCTTTAAACATTAAGCGGCAATTAAGAGTGC 83

```

```

OY 29 AlValaInserMetThrAsnThrArgThrAspValGluAlaThrValAsnGlnIle 48
Db 84 GTCAATCAAGCATGACACACGAAACACATACCTTGAAGACCGTCCGCAATC 143
OY 49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
Db 144 AACGACTCGCGAAGCAGATGTCAAATCGTCGCGCTCGCTGATGAAAGCGCT 203
OY 69 AlagluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHis 88
Db 204 GCCGAGCCCATTCACAGATCAAAAGCGATATCCATCCCTCTGTCGATGAAATTCAT 263
OY 89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysAlaArgAspLys 108
Db 264 TTCACATAATAATTGGCATTAAGAGCATGAAAGCGGAGCCATTAATATCCCATCAT 323
OY 109 ProGlyAsnIleGlyAsnGlnGluArgIleArgMetValValAspCysAlaArgAspLys 128
Db 324 CCGGGTACATCGCGCGCGCGGAAAGTTGAAGCGCTGTCACACGCAAGAGAAAG 383
OY 129 AsnIleProIleArgIleGlyValAlaAsnAlaGlySerLeuGluLysAspLeuGlnLys 148
Db 384 GGCATTCGATCGGATTCGCGCTCATGAGCTCTCTGCAAAACGAAATCCTTGAGAG 443
OY 149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
Db 444 TACGGCTATCCGACACACACGACGATGTCGAAAGCGCCCTGACACATTAATATCTT 503
OY 169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
Db 504 GAGATCTCGATTTCCACATATCATGTCAGATGAAAGCGCGCTGATATTAACCTCGCG 563
OY 189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
Db 564 ATTGAGGCATATGAAAGACCGCTAAAGCTTCGATTAATCCCTTCATTTAGCATACCC 623
OY 209 GluAlaGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeu 228
Db 624 GAATCGGAAACCTGTTTCCGCTACAGTAAAGCGCGCGCTCGCGCATCTCT 683
OY 229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluIle 248
Db 684 TCAGAAAGCATGCGCAATGATGGGATTTCTTAAGCGCGACCGGTCAGAAAGTA 743
OY 249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
Db 744 AAGTCGCCAGGAGGAGCTCTGAAATCTTCGGGCTCGCTCCAAATGCGCAACATTTGATT 803
OY 269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu 288
Db 804 TCCTCCCGGACTTGGCGCGGATGCAATTCGATTTGATTCGATTCGCAATGCAATTCGAA 863
OY 289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn 308
Db 864 GATTACATCGCGAAATACAGCGCGCATCAAAAGTTGCGTTCGCGTGGCGCTCAAC 923
OY 309 GlyProGlyGluAlaLeuValSerThrLeuGlyValIleArgGlyAsnLys 325
Db 924 GGTCCCGGAGAAAGCGCGCAAGCGGATTCGCGGCGCGCAAGCGCGG 974

```

## RESULT 13

AB081842  
ID AB081842 standard; DNA; 349980 BP.

AC AB081842;

DT 19-NOV-2002 (first entry)

DE Bifidobacterium longum NCC2705 genomic sequence SEQ ID NO:1.

KW Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;  
antidiarrhetic; antibacterial; inhibitor of Salmonella; detection;

```

KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
KW rotavirus; food composition; pharmaceutical composition; gene; ds.
OS Bifidobacterium longum.
XX
XX EP1227152-A1.
XX
XX 31-JUL-2002.
XX
XX 30-JAN-2001; 2001EP-0102050.
XX
XX 30-JAN-2001; 2001EP-0102050.
PR
XX (NEST ) SOC PROD NESTLE SA.
PA
XX WPI; 2002-668397/72.
DR
XX
XX Novel polynucleotide comprising Bifidobacterium genome sequence useful
PT as a probe or primer for detecting and/or identifying Bifidobacterium
PT longum in a biological sample
XX
XX Claim 1: SEQ ID 1; 80pp; English.
XX
XX The present invention describes a polynucleotide (I) comprising a
CC sequence of a Bifidobacterium genome selected from the nucleotide
CC sequences given in AB081842 and AB081843, or a sequence exhibiting at
CC least 90% identity or which hybridises with the sequences given in
CC AB081842 and AB081843. Also described is a polynucleotide (II) encoding
CC a fusion protein, comprising a sequence selected from 1097 sequences
CC given in ABP65258 to ABP65354 ligated in frame to a polynucleotide
CC encoding a heterologous polypeptide. (I) has antidiarrhetic and
CC antibacterial activities, and can be used as an inhibitor of Salmonella.
CC (I) (which is a probe) is useful for the detection and/or identification
CC of Bifidobacterium longum in a biological sample. A carrier containing
CC the lactic acid bacterium Bifidobacterium longum NCC2705 (NCIM I-2618)
CC can be used for preventing and/or treating diarrhoea brought about by
CC pathogenic bacteria and/or rotavirus. The carrier is a food composition
CC selected from milk, yogurt, curd, cheese, fermented milks, milk based
CC fermented products, ice-creams, fermented cereal based products, milk
CC based powders, infant formula, pet food or a pharmaceutical composition
CC selected from tablets, liquid bacterial suspensions, dried oral
CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.
CC (I) is useful in DNA arrays or chips to carry out analysis of the
CC expression of the Bifidobacterium gene. AB081844 to AB081850 represent
CC Bifidobacterium related nucleotide sequences given in the Sequence
CC listing from the present invention but not mentioned further within the
CC specification.
CC N.B. The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied by the
CC European Patent Office.
XX
XX
SQ Sequence 349980 BP; 72540 A; 102738 C; 103221 G; 71481 T; 0 other;

```

## Alignment Scores:

Pred. No.:	1.95e-71	Length:	349980
Score:	814.50	Matches:	167
Percent Similarity:	65.65%	Conservative:	70
Best Local Similarity:	46.26%	Mismatches:	111
Query Match:	43.65%	Indels:	13
DB:	24	Gaps:	4

US-09-921-992-78 (1-372) x AB081842 (1-349980)

```

OY 3 AsnGlnAlaProIleGln---ArgArgLysSerThrArgIleTyrValGlyAsnValPro 21
Db 118761 AGCGAATCGCCCTTGACACCGCGCGCAATCCCGCGCATCATGATGCGGTCGCG 118820
OY 22 IleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAspVal 41
Db 118821 GTGGCGGAGAGCGCCCATCTCGATCGATGACCAACAGCGTCACGCGCAACGTA 118880
OY 42 GluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgVal 61

```



```

QY 129 AsnIleProIleArgIleGlyValAlaAsnAlaGlySerLeuGluLysAspLeuGlnLys 148
DB 304 AATATTCCAAATCCGATTCGGCTTAATGCTGATTTAGAAAAGAAAATATTATTCAAA 363
QY 149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
DB 364 TACGGTTACCCATCGCTGAGGAATGGTAGAAATGCTGATCATATTAATTAATCTC 423
QY 169 AspArgLeuAsnPheAspGlnPheGlyValSerValLysAlaSerAspValPheLeuAla 188
DB 424 GAAAGTTTATGATTTTATGATATCATTTCTTTGAAAGCTCTGATGTAATTTAGCA 483
QY 189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
DB 484 ATTGAAGCTTATGATTAAGTAGTCAGCCATTATATTCCTCATCATCTTGGAATTC 543
QY 209 GluAlaGlyValAlaArgSerGlyValAlaLysSerAlaIleGlyLeuGlyLeuLeu 228
DB 544 GAAATCGTACCGCAAGTTTGTGAGGAATTAAGAGTCTGCTGTTAGAGAGCATATC 603
QY 229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle 248
DB 604 AGTTGGCATTTGGAATATACATTAACGGTATCTTGAGTCTGATCCTGTGGAGAGATA 663
QY 249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
DB 664 AAGTGGCCCGCGAGAGTTTAAATCGTTGCTTCGCGAATGCGCGCATCTTATC 723
QY 269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu 288
DB 724 TCCGCGCCCTACTTCGCTCGCAATAGCATTTGATTTGCTGATCGCTAATGAAGTCAA 783
QY 289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValAlaAsn 308
DB 784 AATTACATAGCAACGATTAAAGCTCGATTAAAGTACCCGCTGCTGCTGCGTCAAC 843
QY 309 GlyProGlyGluAlaLeuValSerThrLeuGlyValAlaThrGlyAsnLysLysSerGly 328
DB 844 GGTCCCGAGAAAGCTCCGGAACGATATCGGAATTCGTGTCATTAATGAGAGGCTT 903
QY 329 LeuTyrGluAspGly-----ValArgLys 336
DB 904 CTTTATGACATGTTAAATCATCCGCAAA 933

```

RESULT 15  
AA14017  
ID AA14017 standard; DNA: 1180 BP.

AC AA14017;  
DT 31-MAR-1999 (first entry)  
XX  
DE H. pylori GHPO 76 gene.  
XX  
KW GHPO protein: Helicobacter infection; gastroduodenal disease; gastritis;  
KM peptic ulcer disease; ss.  
XX  
OS Helicobacter pylori.  
FH  
FT Key Location/Qualifiers  
FT CDS 51..1130  
FT /tag= a

MO9843478-A1.  
XX  
PD 08-OCT-1998.  
XX  
PE 01-APR-1998: 98MO-US06371.  
XX  
PR 29-JUL-1997: 97US-0902615.  
PR 01-APR-1997: 97US-0833457.  
PR 24-JUN-1997: 97US-0881227.

```

XX (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
DR WPI: 1998-542293/46.
DR P-PSDB: AAW98298.
XX
XX New isolated Helicobacter polynucleotides - used to develop products
XX for the diagnosis, prevention and treatment of Helicobacter
XX infections and gastrointestinal diseases
XX
XX Claim 1; Page 372-374; 2054pp: English.
XX
XX This sequence represents a polynucleotide of the invention. It was
XX isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
XX The polypeptides can be used for preventing or treating Helicobacter
XX infections, and gastroduodenal diseases associated with these
XX infections, including acute, chronic, and atrophic gastritis, and peptic
XX ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
XX for the production of antibodies. The products can also be used for
XX detection and diagnosis.
XX
XX Sequence 1180 BP; 380 A; 206 C; 289 G; 305 T; 0 other:
XX
XX Alignment Scores:
XX Pred. No.: 1,21e-72 Length: 1180
XX Score: 794.00 Matches: 162
XX Percent Similarity: 66.38% Conservative: 73
XX Best Local Similarity: 45.76% Mismatches: 117
XX Query Match: 42.55% Indels: 2
XX DB: 19 Gaps: 2
XX
XX US-09-921-992-78 (1-372) x AA14017 (1-1180)
QY 9 ArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaProIle 28
DB 63 AGAGTTAAAGACCAAGAAATTTTATTCGTCGCGCGCATATGAGGCGGATGCTCCCA 122
QY 29 AlaValGlnSerMetThrAsnThrArgThrAspValGluAlaThrValAsnGlnIle 48
DB 123 ACCACGCAAAAGCATGACCTTTAGCAAAACCGCTGATATGAAGCACTAAATCAATT 182
QY 49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
DB 183 GACAGACTCAAAATCGCCGCGGCGCATTTAGTACGGTGGCGGTGATGAATGAAGAGAC 242
QY 69 AlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHis 88
DB 243 GCTCTAGCCCTTAAAGAAATTAAGTAAGTGCCTTTCCTTTAATGCTGATATTCAT 302
QY 89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
DB 303 TTCATTTATTAATTCGCTC--ATTGCCCTCAAAACCGTGATGCGATCAGGATTAAC 359
QY 109 ProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLys 128
DB 360 CCGGAAACATCGGCTCAAAAGAGATCAAAAGCGGTGTGATGCTTGTAAAGAAAAA 419
QY 129 AsnIleProIleArgIleGlyValAlaAsnAlaGlySerLeuGluLysAspLeuGlnLys 148
DB 420 AACATTCCTATTAAGAAATGCGGTGAATGCTGGAGTTTAGAAAAGCAATTGATCAAAA 479
QY 149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
DB 480 TACGGA---CCACCCCAAAAGGCATGTGAAAAGCCTTTGATTAACGCCAAACTTTA 536
QY 169 AspArgLeuAsnPheAspGlnPheGlyValSerValLysAlaSerAspValPheLeuAla 188
DB 537 GAAAGTTTGAATTTTACCAATTTTAAGATTTCTTTAAAGACGACGATGATTCGACCC 596
QY 189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr

```

```
Db      597  :::::::::::::::  ::  ||  :::::::::::::::
      209  GlnAlaGlyAlaArgSerGlyAlaValSerAlaIleGlyLeuGlyLeuLeu 228
      657  GAGCGGGGAATCTTTAGCTCCAGTATCAATCCCTATAGGCTTAGGGGCTTTA 716
      229  SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluIle 248
      717  ATGAGGGGCATTTGGGATACATCGCGTATCATCACAGGGGCAATTAGAAATGAATC 776
      249  LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
      777  AAAGTGCCAGAGCAATTTTACGCCATAGCGGCGGTTGAAAGAAGAGGATTAATTGKAT 836
      269  AlaCysProThrCysSerArgGluGluPheAspValIleGlyThrValAlaAsnAlaLeuGlu 288
      837  TCTTGCCCTTGGCGGCGCATTTGAACCAATTTAGTGATATGGCATCAGGTAGAA 896
      289  GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValAlaAsn 308
      897  AAACGCTTAAGCACATCAAAACCCCTTACACATTACGTGATGGTGGTGGTGAAT 956
      309  GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyAsnLysLysSerGly 328
      957  GCTTTGGGTGAAGCCAGCATGACAGCATGGCGATCGCTTTGGGAATCGCAGCGGTTG 1016
      329  LeuTyrGluAspGlyValArgLysAspArgLeuAspAsnAspMetIleAspGlnLeu 348
      1017  ATCATTAAGAAGGCTAAAGTCTTCACAACCTGCTGAAGGATTTATTGAAACTTTT 1076
      349  GluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAla 362
      1077  GTGATAGAGAGTGAAATTTAGCTAAGAGAGAGAAAAAGT 1118
```

Search completed: August 15, 2003, 01:38:42  
Job time : 3223 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 14, 2003, 23:20:23 ; Search time 83 seconds  
(without alignments)  
1978.247 Million cell updates/sec

Title: US-09-921-992-78

Perfect score: 1866  
Sequence: 1 MINQAPIQRRKSTRIVGVN.....RAKASQLEARRIDVQOVER 372

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL-frame\_p2n.model -DEV-x1h  
-Q-/cgn2\_1/USPTO.spool/US09921992/rnat\_12082003\_094913\_17389/app.query.fasta\_1.519  
-DB-issued\_Patents\_NA\_QFMT-fastp -SUFFIX-rni -MINMATCH-0 -LOOPEXT-0  
-LOOPEXT-0 -UNITS-bits -START-1 -END-1 -MAPRIB-biosum62 -TRANS-human40.cdi  
-LIST-45 -DOCALLIGN-200 -THR SCORE-pct -THR MAX-100 -THR MIN-0 -ALIGN-15  
-MODE-LOCAL -OUTPM-pto -NORM-ext -HEAPSIZE-500 -MINLEN-0 -MAXLEN-2000000000  
-USER-US09921992/cgcn\_1\_1\_56-rtinat\_12082003\_094913\_17389 -NCPU-6 -ICPU-3  
-NO\_MAP -LARGEODDERY -NEG\_SCORES-0 -WAIT -DSPBLOCK-100 -LONRLOC  
-DEV\_TIMEOUT-120 -WARN\_TIMEOUT-30 -THREDS-1 -XGAPOP-10 -XGAPEXT-0.5 -FGAPOP-6  
-FGAPEXT-7 -YGAPOP-10 -YGAPEXT-0.5 -DELOP-6 -DELEXT-7

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/prodata/2/lna/5A.COMB.seq:\*  
2: /cgn2\_6/prodata/2/lna/5B.COMB.seq:\*  
3: /cgn2\_6/prodata/2/lna/5A.COMB.seq:\*  
4: /cgn2\_6/prodata/2/lna/5B.COMB.seq:\*  
5: /cgn2\_6/prodata/2/lna/6B.COMB.seq:\*  
6: /cgn2\_6/prodata/2/lna/Dackfilest1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1691	90.6	1010	US-08-827-190-8
2	1691	90.6	1010	US-09-170-187-8
3	1624.5	87.1	1345	US-08-827-190-4
4	1624.5	87.1	1345	US-09-170-187-4
5	1603.5	85.9	1830121	US-09-557-884-1
6	1603.5	85.9	1830121	US-09-643-990A-1
7	1593.5	85.4	1079	US-08-827-190-7
8	1593.5	85.4	1079	US-09-170-187-7
9	1372.5	73.6	1644	US-09-252-991A-6420
10	1372.5	73.6	2202	US-09-252-991A-6131
11	1363.5	73.1	2088	US-09-252-991A-6339
12	1353	72.5	886	US-08-827-190-10

13	1353	72.5	886	US-09-170-187-10	Sequence 10, Appl
14	1195	64.0	1137	US-09-328-3780	Sequence 3780, Ap
15	812	43.5	886	US-08-827-190-9	Sequence 9, Appl1
16	812	43.5	886	US-09-170-187-9	Sequence 9, Appl1
17	764	40.9	555	US-09-252-991A-6251	Sequence 6251, Ap
18	751.5	40.3	4403765	US-09-103-840A-1	Sequence 2, Appl1
19	751.5	40.3	4411529	US-09-252-991A-6209	Sequence 6209, Ap
20	607.5	32.6	543	US-09-221-017B-309	Sequence 1, Appl1
21	470.5	25.2	1230025	US-09-252-991A-1267	Sequence 784, App
22	288.5	15.5	1083	US-08-916-421B-1	Sequence 1, Appl1
23	112.5	6.0	3177	US-08-975-762-39	Sequence 39, Appl
24	111.5	6.0	3988	US-09-066-046-5	Sequence 5, Appl1
25	111.5	5.9	1369	US-09-221-017B-1017	Sequence 1017, Ap
26	106.5	5.7	1854	US-09-252-991A-14350	Sequence 14350, A
27	106.5	5.7	2943	US-09-252-991A-1267	Sequence 14267, A
28	106	5.7	2259	US-09-252-991A-784	Sequence 784, App
29	106	5.7	166976	US-08-916-421B-1	Sequence 1, Appl1
30	105.5	5.7	2129	US-08-975-762-39	Sequence 39, Appl
31	105.5	5.7	2129	US-09-295-028-39	Sequence 39, Appl
32	105.5	5.7	2129	US-09-106-582-39	Sequence 39, Appl
33	103.5	5.5	2214	US-07-985-458-1	Sequence 1, Appl1
34	101.5	5.4	1694	US-09-136-421-11	Sequence 11, Appl
35	101.5	5.4	2989	US-07-927-851-1	Sequence 1, Appl1
36	101.5	5.4	2989	US-08-453-323-1	Sequence 1, Appl1
37	101.5	5.4	2989	US-08-440-520-1	Sequence 1, Appl1
38	101	5.4	1149	US-09-252-991A-13125	Sequence 13125, A
39	101	5.4	1398	US-09-252-991A-12816	Sequence 12816, A
40	101	5.4	1572	US-09-612-964-1	Sequence 1, Appl1
41	100.5	5.4	1663	US-09-328-352-3604	Sequence 3604, Ap
42	100.5	5.4	2010	US-09-039-859-1	Sequence 1, Appl1
43	100.5	5.4	11601	US-08-222-617A-3	Sequence 3, Appl1
44	100.5	5.4	11601	US-08-222-617A-24	Sequence 24, Appl1
45	99.5	5.3	3414	US-09-252-991A-8258	Sequence 8258, Ap

#### ALIGNMENTS

RESULT 1  
US-08-827-190-8  
Sequence 8, Application US/08827190  
Patent No. 5858367  
GENERAL INFORMATION:  
APPLICANT: Rather, Philip N.  
TITLE OF INVENTION: Methods for Screening For Antimicrobials  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESS: Medien & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/827,190  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Peter G.  
REGISTRATION NUMBER: 32,837  
REFERENCE/DOCKET NUMBER: CASE-02443  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1010 base pairs  
TYPE: nucleic acid

STRANDEDNESS: double  
 TOPOLOGY: unknown  
 MOLECULE TYPE: DNA (genomic)  
 us-08-827-190-8

## Alignment Scores:

Pred. No.:	1.43e-185	Length:	1010
Score:	1691.00	Matches:	336
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	90.62%	Indels:	0
		Gaps:	0

US-09-921-992-78 (1-372) x US-08-827-190-8 (1-1010)

```

QY 1 MethIsaNgInAlaProIleGInARgLySeRThrArGLeTyRvAlGlyAsnVal 20
DB 1 ATGCATTAACCGAGGTCCCAATTCACAGTGAATAATCAACACGCTATTTCGTGGGAATGTG 60
QY 21 ProIleGlyAspGlyAlaProIleAlaValGInSeRmetThraSntThrArGThrThAsp 40
DB 61 CCGATTGGCGATGGTGTCCCATCCCGTACAGTCCATGACCAATACGCGTACGACAGAC 120
QY 41 ValGInuAlaThrValAsnGInuIleLysAlaLeuGInuArGyAlGlyAlaAspIleValArG 60
DB 121 GTCGACGCAACGCTCAATCAATCAAGGCGCTGGAGCGCTGGCGCTGATATCGTCCT 180
QY 61 ValSerValProThrMetAspAlaIaGInuAlaPheLysLeuIleLysGInuValAsn 80
DB 181 GTATCCGTACCGACGATGAGCGACGCAAGCGCTTCAACATCAACACACAGGATTAC 240
QY 81 ValProLeuValAlaAspIleHisPheAspTyraRgIleAlaLeuLysValaIaGInuTyR 100
DB 241 GTGGCGGTGGGTGACATCACTTGACATCGCATTCGCGTGAAGTACGGAATAC 300
QY 101 GlyValAspCysLeuAlaGlyIleAsnProGlyAsnIleGlyAsnGInuArGyIleArgMet 120
DB 301 GGCCTGCTGCTGCGCGATTAACCCCTGGCAATATCGGTATGGAAGACGATTCCGCTAG 360
QY 121 ValValAspCysAlaArgAspIleAsnIleProIleArgIleGlyValaIaAsnAlaGlySer 140
DB 361 GTGGTGTGCTGCGCGCGATTAACCACTTCGATTCCTATTGGCGTTAACGCGGATCG 420
QY 141 LeuGInuLysAspLeuGInuLysTyRgIyLupProThrProGInuAlaLeuLeuGInuSer 160
DB 421 CTGGAAAAAGATCTGCAAGAAAGATGCGAACCGACGCGCGCTTCTCTGGAATCT 480
QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGInuPheLysValSerVal 180
DB 481 GCCATGCGTCATGTGATCATCTGCATCGCTGAACCTTGATCAAGTCAAGCTCAGCGTG 540
QY 181 LysAlaSerAspValPheLeuAlaValGInuSerTyraRgLeuLeuAlaLysGInuIleAsp 200
DB 541 AAAGGCTGTGACGCTTCTCGCTGCTGAGTCTTTCGTTGCTGGCAAAACAGATCAT 600
QY 201 GInProLeuHisLeuGlyIleThrgInuAlaGlyIleAlaArgSerGlyAlaValLysSer 220
DB 601 CAGCGGTGCAATCTGGGGATCACCGAACCGCGTGGTGGCGGAGGGGGCGAGTAAATCC 660
QY 221 AlaIleGlyLeuGlyLeuLeuLeuSerGInuGlyIleGlyAspThrLeuArGyAlaSerLeu 240
DB 661 GCCATTGGTTTAGTCTGCTGCTGCTGCTGAAGGCAATCGCGACGCTCGCGCTATCGCTG 720
QY 241 AlaAlaAspProValGInuGInuIleLysValGlyPheAspIleLeuLysSerLeuArGyIle 260
DB 721 GCGGCGGATCGCGGTGAGAGATCAAGTCCGTTTCGATATTGAAATCCGCTCGGTATC 780
QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGInuPheAspVal 280
DB 781 CGTTTCGAGAGGATCAACTTCATCCCTGCGCGACCTCTCCGCGAGGAATTTGATGTT 840
QY 841 IleGlyThrValAsnAlaLeuGInuArGyLeuGInuAspIleIleThrProMetAspVal 300
  
```

```

DB 841 ATCGGTACGGTTAAGCGGTGAGCAACGCTGGAGATATCATCTCCGANGAGCCTT 900
QY 301 SerIleIleGlyCysValValaSnGlyProGlyGInuAlaLeuValSerThraGlyVal 320
DB 901 TCGATTATCGCGCTGCGGTGGAATGGCCAGGTGAGCGCTGCTTCTACACTCGCGCTC 960
QY 321 ThrGlyLysAnLysLysSerGlyLeuTyRgInuAspGlyValaArGlys 336
DB 961 ACCGCGGCAACAAGAAACCGGCTCTATGAAGATGCGTGGCGCAA 1008
  
```

## RESULT 2

US-09-170-187-8  
 Sequence 8, Application US/09170187

Patent No. 6383745

GENERAL INFORMATION:

APPLICANT: Rather, Philip N.  
 TITLE OF INVENTION: Methods For Screening For Antimicrobials

TITLE OF INVENTION: Utilizing aarc And Compositions thereof

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Medlen & Carroll, LLP  
 STREET: 220 Montgomery Street, Suite 2200  
 CITY: San Francisco

STATE: California  
 COUNTRY: United States of America

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentln Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/170,187

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/827,190

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Carroll, Peter G.

REGISTRATION NUMBER: 32,837

REFERENCE/DOCKET NUMBER: CASE-02443

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 1010 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)

US-09-170-187-8

## Alignment Scores:

Pred. No.:	1.43e-185	Length:	1010
Score:	1691.00	Matches:	336
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	90.62%	Indels:	0
		Gaps:	0

US-09-921-992-78 (1-372) x US-09-170-187-8 (1-1010)

```

QY 1 MethIsaNgInAlaProIleGInARgLySeRThrArGLeTyRvAlGlyAsnVal 20
DB 1 ATGCATTAACCGAGGTCCCAATTCACAGTGAATAATCAACACGCTATTTCGTGGGAATGTG 60
QY 21 ProIleGlyAspGlyAlaProIleAlaValGInSeRmetThraSntThrArGThrThAsp 40
DB 61 CCGATTGGCGATGGTGTCCCATCCCGTACAGTCCATGACCAATACGCGTACGACAGAC 120
QY 41 ValGInuAlaThrValAsnGInuIleLysAlaLeuGInuArGyValGlyAlaAspIleValArG 60
  
```



```

Db      121 GTGAAACCAAGCGCATCAAAATCAGCGCTGGAAGCGCTGGCGCTGATCGTCCT
Qy      61 ValSerValProThrMetAspAlaAlaGluAlaPheLeuLeuLeuSgInGlnValAsn
Db      181 GTATCCGTACCGACGATGAGCGCGGCAAGACGCTCAACTCAATCAACAGCGGTTAAC
Qy      81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr
Db      241 GTCCCGCTGGCTGCATCCACTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC
Qy      101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyValAsnGluAlaGlySer
Db      301 GGGCGATGTGTGGCTATTAACTCGCAATATGCTAATGAAGAGCGCTATTCATG
Qy      121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAlaAsnAlaGlySer
Db      361 GTGGTGACGTGCGCGCGGCGATTAATAACATTCGATTCGATTCGATTCGATTCGATTC
Qy      141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuGluSer
Db      421 CTGGAAAGATCTGCAGAAAGATGGGCAACGACGCGCGGCGGCTGGATCG
Qy      161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal
Db      481 GCCATCGTCATGTTGATCATCTCGATCGCTGAACCTTGATCATCAAGTCAAGCAGCTG
Qy      181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp
Db      541 AAGCGCTGACGCTTCCTCGCTGTTGATCTATTCGTTGCTGGCAAAACGATCGAT
Qy      201 GlnProLeuHisLeuGlyIleThrGluAlaGlyValAlaArgSerGlyValAlaLysSer
Db      601 CAGCCCTTCATCTGGGGAATCACCCAGCGGCTGGGCGGAGCGGCGGAGTAAATCC
Qy      221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu
Db      661 GCCATTGGTTAGTGTGCTGCTGCTGTAAGGCAATCGGACACGCTGGCGCTGCTG
Qy      241 AlaAlaAspProValGluGluLeuLysValGlyPheAspIleLeuLysSerLeuArgIle
Db      721 GCGGCGGATCCGCTGCAAGATCAAAAGTGGCTTCGATTTGAAATCGCTGCGTATC
Qy      261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnIlePheAspVal
Db      781 CTTGCGAGGAGTCACTTCATCGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG
Qy      281 IleGlyThrValAlaAsnAlaLeuGlnGluArgLeuGluAspIleIleThrProMetAspVal
Db      841 ATCGGTACGCTTAAGCGCTGAGACACGCTGGAAGATATCATCATCTCCGATGAGCGTT
Qy      301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal
Db      901 TCGATTATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
Qy      321 ThrGlyGlyAsnLysSerGlyLeuTyrGluAspGlyValAlaLys 336
Db      961 ACCGCGGCAACAGAAAGAGCGCTCTATGAAGATGGCGTGGCAAA 1008

```

```

? COUNTRY: United States of America
? ZIP: 94104
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent In Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/827,190
? FILING DATE:
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Carroll, Peter G.
? REGISTRATION NUMBER: 32,837
? REFERENCE/DOCKET NUMBER: CASE-02443
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (415) 705-8410
? TELEFAX: (415) 397-8338
? INFORMATION FOR SEQ ID NO: 4:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1345 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 175..1272
? OTHER INFORMATION: /gene="aarc"
? OTHER INFORMATION: /note="Similar to E. coli Gcpe protein listed by GenBank."
US-08-827-190-4

Alignment Scores:
Pred. No.: 1 02e-177 Length: 1345
Score: 1624.50 Matches: 325
Percent Similarity: 94.09% Conservative: 25
Best Local Similarity: 87.37% Mismatches: 21
Query Match: 87.06% Indels: 2
DB: 2 Gaps: 1

US-09-921-992-78 (1-372) x US-08-827-190-4 (1-1345)
Qy      1 MethHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
Db      175 ATGCATTAATGAATCAACCGTAAAGAGCTAATATCCACCGAATTTATGTAAGTACGTG 234
Qy      21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
Db      235 CCTATTGGCGAGTGTGCTCCCATTCCTGTCATCTATGAGCAATATACCGCGACGAGGAT 294
Qy      41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
Db      295 GTTGAAGCATCTGTCCGCAATCCATCACTGAGCGGTGAGGTGATGATCGTCCGC 354
Qy      61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
Db      355 GTGTCTGTCTACAGATGATGACAGACAGAGCTTTAATTAATTAATTAACACGCGCTGAT 414
Qy      81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
Db      415 GTGCCATTGTGGCGGATATTCACCTTGTACACCGTATCCGAGTAAAGTGGCTGAATAT 474
Qy      101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyValAsnGluArgIleArgMet 120
Db      475 GTGTGTGACCTCCACAGATTAATACCGAGTAAATTCGCGACGTGAAGACGTAATTCGCCAA 534
Qy      121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAlaAsnAlaGlySer 140
Db      535 GTCGTTGATGTGCTGCTGATCATCAACATTCCTATCCGTATAGGGGTCAATGGCGGCA 594
Qy      141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLysGluSer 160
Db      595 CTGGAAGAAAGATATCCAGAAAGATACGGTGGCAACCTCTGAAGCATGTGTTGAATCA 654

```

```

RESULT 3
US-08-827-190-4
Sequence 4, Application US/08827190
Patent No. 5858367
GENERAL INFORMATION:
APPLICANT: Rather, Philip N.
TITLE OF INVENTION: Methods For Screening For Antimicrobials
TITLE OF INVENTION: Utilizing aarc And Compositions Thereof
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
SERIAL: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California

```

```

OY 161 AlameTarGhiSVAlaSPHISLeuSPARLeuSPnPhenSPGlnPheLYSValSerVal 180
    |||||||
Db 655 GCATATGGACATGTTGATATCTTGACAGCGCTGATATTCATCATGTTCAAGGTCAGGTT 714
OY 181 LysAlaSerSPValPheLeuAlaValGluSerTyrArgLeuAlaLYSGlnIleASP 200
    |||||||
Db 715 AAAGCGTGGATGCTTCTTCCGCGTGGCTCTTATCGTTATGCGCAAAAATGAT 774
OY 201 GlnProLeuHISLeuGlyIleThrGlnAlaGlyLYSAlaArgSerGlyAlaValLYSSer 220
    |||||||
Db 775 CAACCACTGCACCTCGGTATATACAGAACGCGTGGGCTGCTCTGCTTCAAGTCAATCA 834
OY 221 AlaIleGlyLeuGlyLeuGlyLeuSerGlyIleGlyAlaSPThrLeuArgValSerLeu 240
    |||||||
Db 835 GCAATGGCTGTGGTATATGTTGGCTGAGAGTATCGCGCATACGTTACGATCTCACTC 894
OY 241 AlaAlaSPProValGluGluIleLYSValGlyPheASPLeuLYSLeuArgIle 260
    |||||||
Db 895 GCGGCAGATCCTGTTGAGGAAGTGAAGTGGTTTGAATTTCAAAATCGTTACGATC 954
OY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheASPVal 280
    |||||||
Db 955 CGCTCAGCGTGCATCACTTATATGCTTGCCCAACCTGTTCAAGCAAGAAATTTGATGTG 1014
OY 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluSPLeuIleThrProMetLeuSPVal 300
    |||||||
Db 1015 ATTGAGTACGTAATATGCTTTGGAGACAGCGCTCGAAGATATATACACCGCATGATGTC 1074
OY 301 SerIleIleGlyCysValValAlaSPGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
    |||||||
Db 1075 TCTATATATGTTGGTGTGATGTAATGGCCGGGTGAAGCCGAGGTTTCTACTTTAGGTGTG 1134
OY 321 ThrGlyGlyAsnLYSLeuSerGlyLeuTyrGluSPGlyValArg--LYSAspArgLeu 339
    |||||||
Db 1135 GCTGCGCGGAAACCAAAAGTGTTCTTCAATGAAGTGGCTTGGCAAAAAGAGCGTTT 1194
OY 340 ASPAsnAspMetIleASPGLnLeuGluAlaArgIleArgAlaLYSAlaSerGlnLeu 359
    |||||||
Db 1195 GATATGACAATATATATGATCAGCTGAGCGGAAATTCGCGCAAAAGCAGCAATGCTT 1254
OY 360 ASPGluAlaArgArgIleASPValGlnGlnValGlu 371
    |||||||
Db 1255 GATGAAT--AACCGTATAAAGATTAACCAAGTCGAA 1289

```

RESULT 4  
US-09-170-187-4

Sequence 4, Application us/09170187

Patent No. 6383745

GENERAL INFORMATION:

APPLICANT: Rather, Philip N.

TITLE OF INVENTION: Methods For Screening For Antimicrobials

TITLE OF INVENTION: Utilizing aarc And Compositions Thereof

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSER: Medlen & Carroll, LLP

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: California

COUNTRY: United States of America

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/170,187

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/827,190

FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Peter G.  
REGISTRATION NUMBER: 32,837  
REFERENCE/DOCKET NUMBER: CASE-02443  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1345 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 175..1272  
OTHER INFORMATION: /gene="aarc"  
OTHER INFORMATION: /note="Similar to E. coli Gcpe protein listed by GenBank."

US-09-170-187-4

Alignment Scores:

Pred. No.:	Length:	Matches:
1.02e-177	1345	325
Score: 1624.50	Conservative: 25	
Percent Similarity: 94.09%	Mismatches: 21	
Best Local Similarity: 87.37%	Indels: 2	
Query Match: 87.06%	Gaps: 1	

US-09-921-992-78 (1-372) x US-09-170-187-4 (1-1345)

```

OY 1 MethISnGlnAlaProIleGlnArgArgSerThrArgIleTyrValGlyAsnVal 20
    |||||||
Db 175 ATGATATATGATATACCGATTAAGAGAGTAAATCCACCCGAATTTATGATAGTAAGCTG 234
OY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrASP 40
    |||||||
Db 235 CCTATGGCGATGGTGGTCCCATTCGCTGCTCAATCTATGACGAATACGCGACGACGAT 294
OY 41 ValGluAlaThrValAsnGlnIleLYSAlaLeuGluArgValGlyAlaASPLeuValArg 60
    |||||||
Db 295 GTTGAGCCACTGTCGCGCAATCCAACTTACGCTGAGGCTGATGATGCTCCG 354
OY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLYSLeuIleLYSGlnGlnValAsn 80
    |||||||
Db 355 GTGCTGCTCTCATGATGATGATGACGACAAACCTTTAAATTAATTAAGACGCGGTGAT 414
OY 81 ValProLeuValAlaASPLeuIleHisPheASPThrArgIleAlaLeuLYSValAlaGluTyr 100
    |||||||
Db 415 GTGCCATGTGTTGGGATATTCACCTTGACTACCGTATCCGAGTGAAGAGTGGTGATAT 474
OY 101 GlyValaSPCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
    |||||||
Db 475 GGTGTGACTGCTCAGATTAATTAACCCAGATTAATATCGCAGTGAAGACGATTCGCCAA 534
OY 141 LeuGluLYSAspLeuGlnGluLYSTyrGlyGluProThrProGlnAlaLeuLeuGlnSer 160
    |||||||
Db 535 GTGCAATGACATGTTGATATCTTGACAGCGTGAATTTGATCACTTAAGCTCAGTGT 714
OY 161 AlameTarGhiSVAlaSPHISLeuSPARLeuSPnPhenSPGlnPheLYSValSerVal 180
    |||||||
Db 655 GCATATGGACATGTTGATATCTTGACAGCGCTGATATTCATCATGTTCAAGGTCAGGTT 714
OY 181 LysAlaSerSPValPheLeuAlaValGluSerTyrArgLeuAlaLYSGlnIleASP 200
    |||||||
Db 715 AAAGCGTGGATGCTTCTTCCGCGTGGCTCTTATCGTTATGCGCAAAAATGAT 774
OY 201 GlnProLeuHISLeuGlyIleThrGlnAlaGlyLYSAlaArgSerGlyAlaValLYSSer 220
    |||||||

```

```

Db 775 CAACCACTTCACTCGGTATTACAGAACGGGTGGGCTCGTCTGTTCAAGTAATCA 834
Qy 221 AAlleGlyLeuGlyLeuLeuSerGlyLeuGlyLeuGlyLeuGlyLeuGlyLeu 240
Db 835 GCAATGTGCTGTGATGATGTTTGGCGAAGTATCGCGATACGTTACGATCTCATC 894
Qy 241 AAlaAAspProValGluGluLeuValGlyLeuValGlyLeuValGlyLeuVal 260
Db 895 GCGGCAATCCTGTTGAGGAGAGTAAGTGGCTTTGATATTCAAAATCGTTACGATC 954
Qy 261 ArgSerArgGlyLeuPheLeuPheLeuGlyLeuGlyLeuGlyLeuGlyLeu 280
Db 955 CGCTCAGCTGCATCAACTTATGCTTGGCCCAACTGCTCACCACCAAGATTTGATG 1014
Qy 281 lIleGlyThrValAsnAlaLeuGluGlnArgLeuGluGluGluGluGluGlu 300
Db 1015 ATTGTCACGTAATGCTTTGGACACGCGCTCAAAATATTATACCGCGATGATGTC 1074
Qy 301 SerLeuLeuGlyValValAlaGlyLeuGlyLeuGlyLeuGlyLeuGlyLeu 320
Db 1075 TCTATTATGTTGTTGATGTAAGTAAGCGCGGTAACCGCATGTTTACTTTAGTGTG 1134
Qy 321 ThrGlyGlyAsnGlyLeuGlyLeuGlyLeuGlyLeuGlyLeuGlyLeuGly 339
Db 1135 GCTGGCGCGAACAACCAAGTGTGTTCTATGAAATGCGCTTCCCAAAAAAGAGCGTTT 1194
Qy 340 AspaAsnAspMetIleAspGlnLeuGluAlaArgLeuArgLeuArgLeuArg 359
Db 1195 GATTAATGACAAATATTATGATCACTGAGCGCAAAATTCGCGCAAAAGCAGCATCTT 1254
Qy 360 AspGluAlaArgArgGlyLeuValGlnValGln 371
Db 1255 GATGAAT- AACGTATTAAGATTAACCAAGTCGAA 1289

RESULT 5
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Felschmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB186P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs

```

```

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Alignment Scores:
Pred. No.: 1,38e-170 Length: 1830121
Score: 1603.50 Matches: 320
Percent Similarity: 93.66% Conservative: 20
Best Local Similarity: 88.15% Mismatches: 22
Query Match: 85.93% Indels: 1
DB: 4 Gaps: 1

US-09-921-992-78 (1-372) x US-09-557-884-1 (1-1830121)
Qy 4 GlnAlaProIleGlnArgArgGlySerThrArgIleArgValGlyAsnValProIleGly 23
Db 390972 CAGCCACATTTTAAGCGTCGTGATGACACAAAATTTATGTGGAAATGTACCAATTTG 391031
Qy 24 AspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAspValGluAla 43
Db 391032 GGGGATGCGCCTATTGCGCGCAATGACAAATCTCGCACCACTGATGTGGAAGCG 391091
Qy 44 ThrValAsnGlnIleuValAlaLeuGluArgValGlyValAspIleValArgValSerVal 63
Db 391092 ACAGTTCCTAAATTAATATTAAGACGTGTGTGACAGATATTTGCTGTATCTGTT 391151
Qy 64 ProThrMetAspAlaAlaGluAlaPheValLeuIleValGlnGlnValAsnValProLeu 83
Db 391152 CCAACATGATGCTGCGGCAAGCATTTAAACAATTAACACAGATGATGTCGCTC 391211
Qy 84 ValAlaAspIleHisPheAspTyrArgIleAlaLeuValAlaGluTyrGlyValAsp 103
Db 391212 GTACGAGATTAATTAATTCGATTCGATTCGCTTAAACCTCCAGAAATTTGAGAGTGAT 391271
Qy 104 CysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValAsp 123
Db 391272 TGTTTTCGATCAATCTGCAACATTTGTCGTAAGATGCGCTCCGCTGTTGAT 391331
Qy 124 CysAlaArgAspIleAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluVal 143
Db 391332 TGTGCGCGAACAATAATATCCGATTCGATTTGTTAATCCAGGCTCTTTAGAAAA 391391
Qy 144 AspLeuGlnGluIleuValArgIleuProThrProGlnAlaLeuLeuGluSerAlaMetArg 163
Db 391392 GATTTGCAGAAATAATATGCGCAACCAACCCGAAAGCCTGTTAATAATCCGATTCGCT 391451
Qy 164 HisValAspHisLeuAspArgLeuAsnPheAspGlnPheValSerValIleValAsp 183
Db 391452 CATGTGAATAATCTAGCTGCTTAATCTTGACATGTTAAAGTGAAGCCGCTCC 391511
Qy 184 AspValPheLeuAlaValGluSerTyrArgLeuAlaValGlnIleAspGlnProLeu 203
Db 391512 GATGTATTTCTACCGGTTGATCTTATCTGTTACTGCTAAACCAATTAAACAGCCTTA 391571
Qy 204 HisLeuGlyIleThrGluAlaGlyValAlaArgSerGlyValAlaValIleSerAlaIleGly 223
Db 391572 CATTTAGCATTAACAAGCAGGTGGCGACCGCGCTGTGCACTAAATACTGAGTGGGT 391631
Qy 224 LeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAsp 243
Db 391632 TTAGGAATGTTTATTACTGAGGCAATTTGGGATACACTACGCGTCTTTGGCGGAGAT 391691
Qy 244 ProValGluGlnIleuValGlyPheAspIleLeuValSerLeuArgIleArgSerArg 263
Db 391692 CCGTAGAGAAATCAATCGGTTTGAATTTTAAATCTTTACGATGCTTCACAA 391751
Qy 264 GlyIleAsnPheIleAlaCysProThrCysSerArgGlnIlePheAspValIleGlyThr 283
Db 391752 GGAATTAATCTTATCTGTCGCCAACCTTTCTCCGCAAAATTTATGATTCGTTACA 391811
Qy 284 ValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIle 303

```

Db 391812 GTAAATCGGTAGAACACCGCTTGAAGATATTATTCACCAATGATGATCTATATTC 391871  
 Oy 304 GTCysValVala)ansg)lProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGly 323  
 Db 391872 GGTGTGTAGTGAATGTCCTGCGGAGCACTGCTCCGATCTGCGGTACCGGGCGT 391931  
 Oy 324 AsnLysSerGlyLeuTyrGluAspGlyValArg--LysAspArgLeuAspAsn 342  
 Db 391932 AACAAAAACCGCTTATATCTTACGAGGAGACCCAAAAAGAGCGTTTGATACGAA 391991  
 Oy 343 AspMetLeaspGlnLeuGluAlaArgLLeaArgAlaLysAlaSerGlnLeuAspGluAla 362  
 Db 391992 GATATGAGAACCAATTAAGCAAAATTCGTGCAAAAGTCGACGACCAAGATCCAAA 392051  
 Oy 363 ArgArgile 365  
 Db 392052 AACCAAT 392060

## RESULT 6

US-09-643-990A-1  
 : Sequence 1, Application US/09643990A  
 : Patent No. 6528289

## GENERAL INFORMATION:

APPLICANT: Robert D. E. Eischmann  
 Mark D. Adams  
 Owen White  
 Hamilton O. Smith  
 J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS: ADDRESSEE: Human Genome Sciences, Inc.  
 STREET: 9410 Key West Avenue  
 CITY: Rockville,  
 STATE: MD

COUNTRY: USA  
 ZIP: 20850

COMPUTER READABLE FORM: MEDIUM TYPE: 3 1/2 inch diskette

COMPUTER: Dell Pentium

OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA: APPLICATION NUMBER: US/09/643,990A  
 FILING DATE: 23-Aug-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA: APPLICATION NUMBER: 08/487,429  
 FILING DATE: 1995-06-07

APPLICATION NUMBER: 08/426,787

FILING DATE: 1995-04-21

ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302

REFERENCE/DOCKET NUMBER: PB186P1C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-610-5790

TELEFAX: 310-309-6439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1830121 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-643-990A-1

Alignment Scores:

Pred. No.: 1,386,370 Length: 1830121  
 Score: 1603.50 Matches: 320

Percent Similarity: 93.66%  
 Best Local Similarity: 88.15%  
 Query Match: 85.93%  
 DB: 4 Gaps: 1

US-09-921-992-78 (1-372) x US-09-643-990A-1 (1-1830121)

Oy 4 GlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnValProIleGly 23  
 Db 390972 CAGCAACATATTAAAGCGCTGCTGATTCACAAAAATTTATGTGGGAAATGTACCAATTTGT 391031  
 Oy 24 AspGlyLapProIleAlaValGlnSerMetThrAsnThrArgThrAspValGluAla 43  
 Db 391032 GGGATGCGCTTATTCGCGTCAATTCATATGCAAAATCTGACACCTGATGTGGAAGCG 391091  
 Oy 44 ThrValAsnGlnIleLysAlaLeuGluArgValAlaAspIleValArgValSerVal 63  
 Db 391092 ACAGTGTCTCAATTAATCAATTAAGCAAGTGTGTCAGATATGTTCGCTGCTGTT 391151  
 Oy 64 ProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnIleValAsnValProLeu 83  
 Db 391152 CCAACAAATGATGTGCGGAGAACATTTAAACAAATTAACAAACAAATGATGTCCTC 391211  
 Oy 84 ValAlaAspIleHisPheAspTyrArgLLeaAlaLeuLysValAlaGluTyrGlyValAsp 103  
 Db 391212 GTACAGATATTCTATTTCGACTATTCGATTCGCTTAAAGTCGACGATATGAGTGGAT 391271  
 Oy 104 CysLeuArgLLeaAsnProGlyAsnIleGlyAsnGlnGluArgLLeaMetValValAsp 123  
 Db 391272 TGTTCAGTATCAATCCGCGCAACATGTCGTGTAAGATCCGTCGCTGCTGCTGAT 391331  
 Oy 124 CysAlaArgAspLysAsnIleProIleArgLLeaGlyValAsnAlaGlySerLeuGluLys 143  
 Db 391332 TGTGCGGAGACAAAAAATATTCCTGATTCGTTATGTTGTTGAATGCAAGCTCTTTGAAGAAA 391391  
 Oy 144 AspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuGlnSerAlaMetArg 163  
 Db 391392 GATTTCGCAAGAAAAATATGCGCAACCAACGCAAGCTGTTAGATCCGCAATTCGCT 391451  
 Oy 164 HisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSer 183  
 Db 391452 CATGTAGAAATTCAGATCGCTTAATCTTCATCTGTTAAAGTAGGCTAAAGCTCC 391511  
 Oy 184 AspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspIleProLeu 203  
 Db 391512 GATGATTCCTAGCGGCTTGAATCTTATTCGCTTAAAGCAATTAACAGCTTTA 391571  
 Oy 204 HisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGly 223  
 Db 391572 CATTTAGGCATTACAGAACAGTGTGCGCGCTGCTGTCAGTAAATTCGCACTGGGT 391631  
 Oy 224 LeuGlyLeuLeuSerGlnGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAsp 243  
 Db 391632 TTAGGAATGTTATTGAGGAGGCAATGCGATACACTACGCGCTCTTTGGCGGCAAT 391691  
 Oy 244 ProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgLLeaArgSerArg 263  
 Db 391692 CCTGTAGAGGAATCAAGGTGCTTTGATATTGGAATCTTACGATTCGTTCAAGA 391751  
 Oy 264 GlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThr 283  
 Db 391752 GGAATTAACCTTATGCTTGTGCGCAACCTGCTCCCAAGATTTGATGATTCGTTCA 391811  
 Oy 284 ValAsnAlaLeuGlnArgLeuGlnAspIleIleThrProMetAspValSerIleLe 303  
 Db 391812 GTAAATGCGTGAACACACGCTTGAAGATATTATTAACCAATGATGATCTATATTC 391871  
 Oy 304 GlyCysValVala)ansg)lProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGly 323  
 Db 391872 GGTGTGTAGTGAATGTCCTGCGGAGCACTGCTCCGATCTGCGGTACCGGGCGT 391931  
 Oy 324 AsnLysSerGlyLeuTyrGluAspGlyValArg--LysAspArgLeuAspAsn 342

Db 391932 AACAAAAAGCGGTATATCTTGACGAGAACGCCAAAGACGCTTTTGATACGAA 391991  
 QY 343 AspmetileaspGlnleuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAla 362  
 Db 391992 GATATAGTGAACCAATTAGAACAAAAATTCGTGCGAAGTCGACGACAAAGATCCAAAA 392051  
 QY 363 ArgArgIle 365  
 Db 392052 AACGAGATT 392060

## RESULT 7

US-08-827-190-7  
 ; Sequence 7, Application US/08827190  
 ; Patent No. 5858367  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rather, Philip N.  
 ; TITLE OF INVENTION: Methods For Screening For Antimicrobials  
 ; NUMBER OF SEQUENCES: 13  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Medlen & Carroll, LLP  
 ; STREET: 220 Montgomery Street, Suite 2200  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: United States of America  
 ; ZIP: 94104  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; FILING DATE:  
 ; CLASSIFICATION: US/08/827,190  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Carroll, Peter G.  
 ; REGISTRATION NUMBER: 32,837  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 705-8410  
 ; TELEFAX: (415) 397-8338  
 ; INFORMATION FOR SEQ. ID NO.: 7:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1079 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: DNA (genomic)  
 ; US-08-827-190-7

## Alignment Scores:

Pred. No.: 2.76e-174 Length: 1079  
 Score: 1593.50 Matches: 317  
 Percent Similarity: 94.93% Conservative: 20  
 Best Local Similarity: 89.30% Mismatches: 17  
 Query Match: 85.40% Indels: 1  
 Gaps: 1

US-09-921-992-78 (1-372) x US-08-827-190-7 (1-1079)

QY 7 IleglnArgArgLysSerThrArgIleArgValGlyAsnValProIleGlyAspGlyAla 26  
 Db 6 ATTAAGCGTCGATCGAATCAAAATTTATGTGGAAATGTCACATTGGTGGGATCGG 65  
 QY 27 ProIleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsn 46  
 Db 66 CCTATTGCGCGCATCAATGACAAATACGACACACAGTGATGGAAGCGACATTCCT 125  
 QY 47 GlnIleLysAlaLeuGluArgValGlyAlaAspIleValGlyValSerValProThrMet 66  
 Db 126 CAAATTAATCATTAAGAAAGTGTGGTGCAGATATATGTCGTATCTGTCCAAACATG 185

QY 67 AspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAsp 86  
 Db 186 GATGCTCGGAGAGATTTAAACAATTAACAAACAACTGATGTCCGCTCGACGAGAT 245  
 QY 87 IleHisPheAspIleArgIleAlaLeuLysValAlaGluThrGlyValAspCysLeuArg 106  
 Db 246 ATTCATTTGCATGATGATCGGTAAAGTGCAGAAATATGAGATGATGTTCATCGT 305  
 QY 107 IleAsnProGlnAsnIleGlnGlnGluArgIleArgMetValValAspCysAlaArg 126  
 Db 306 ATCAATCCGCGCAACATGTGCTGTAAGATCGCGCTGCTGTGTGATTTGCGCGA 365  
 QY 127 AspLysAsnIleProIleArgIleGlyValAlaAlaGlySerLeuGlnLysAspLeuGln 146  
 Db 366 GACAAAATATTCGATTCGTATGTTGTTAAATGACGAGCTCTTACAAAAAGATTGCCA 425  
 QY 147 GlnLysThrGlyGluProThrProGlnAlaLeuLeuGlnSerLysMetArgHisValAsp 166  
 Db 426 GAAAAATATGCGCAACCAACGCGAAGACCTTGTAGAAATCCGCAATGGCTCATGTAGA 485  
 QY 167 HisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPhe 186  
 Db 486 ATTCATGATCGTCTTAACCTTCAGTTCATTAAAGTACGAGGTAAACCTCCGATGATTC 545  
 QY 187 LeuAlaValGlnSerThrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGly 206  
 Db 546 TTACGCTGTAATCTTATCTGTTACGCTTAAGCAATTAACAGCTTATACATTAGAC 605  
 QY 207 IleThrGlnAlaGlyValAlaArgSerGlyValAlaValLysSerAlaIleGlyLeuGlyLeu 226  
 Db 606 ATTACAGAACGAGTGGCGCACGCGGTGTCAGTAAATTCGCACTGGCTTTAGCAATG 665  
 QY 227 LeuLeuSerGlnGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGlu 246  
 Db 666 TTATTAGCTGAGGCGCATTTGGCATACACACCTCTCTTGGCGCAATCTCGTAGAG 725  
 QY 247 GlnIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyLeuAsn 266  
 Db 726 GAAATCAAAATGCGGTTTATATTTGAAATCTTTACGATTCGTTCAAGAGCAATTAAC 785  
 QY 267 PheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAla 286  
 Db 786 TTTATGCTTGGCCCAACCTGTTCTGCGCAAGATTTGATGATGCGTACATAATGCG 845  
 QY 287 LeuGlnArgLeuGlnAspIleIleThrProMetAspValSerIleIleLysVal 306  
 Db 846 CTGAAACAAAGCCCTTGAAATATTTATACCAATGATGATATATTCGTTGCTGTA 905  
 QY 307 ValAsnGlyProGlyGlnAlaLeuValSerThrLeuGlyValThrGlyLysAsnLys 326  
 Db 906 GTGATGCTGCTGGGAGGCACTGCTCCGATCTCGGCGTAAACGCGGTAAACAAAAA 965  
 QY 327 SerGlyLeuThrGlnAspGlyValArg--LysAspArgLeuAspAsnAspMetIle 345  
 Db 966 AGCGGTATTAATCTTGACGAGAACGCCAAAAAGACCTTTGTAACGAGATATAGTG 1025  
 QY 346 AspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360  
 Db 1026 AACCAATTAGAAAGAAAAATTCGTGCAAAAGTCGACGACCAAGAT 1070

## RESULT 8

US-09-170-187-7  
 ; Sequence 7, Application US/09170187  
 ; Patent No. 6383745  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rather, Philip N.  
 ; TITLE OF INVENTION: Methods For Screening For Antimicrobials  
 ; NUMBER OF SEQUENCES: 13  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Medlen & Carroll, LLP  
 ; STREET: 220 Montgomery Street, Suite 2200  
 ; CITY: San Francisco

STATE: California  
 COUNTRY: United States of America  
 ZIP: 94104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/170,187  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/827,190  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Carroll, Peter G.  
 REGISTRATION NUMBER: 32,837  
 REFERENCE/DOCKET NUMBER: CASE-02443  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 705-8410  
 TELEFAX: (415) 397-8338  
 INFORMATION FOR SEO ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1079 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: unknown  
 MOLECULE TYPE: DNA (genomic)  
 US-09-170-187-7

Alignment Scores:  
 Pred. No.: 2,766-174 Length: 1079  
 Score: 1593.50 Matches: 317  
 Percent Similarity: 94.93% Conservative: 20  
 Best Local Similarity: 89.30% Mismatches: 17  
 Query Match: 85.40% Indels: 1  
 Gaps: 1

US-09-921-992-78 (1-372) x US-09-170-187-7 (1-1079)

QY 7 IleglnarGArglySerThrArgIleTyValGlyAsnValProIleGlyAspVala 26  
 Db 6 ATTAAGCGTCGATCGACAAATAATTATGTGGGAAATGTCACAAATGGTGGGATGCG 65  
 QY 27 ProIleAlaValGlnSerMetThrAsnThrArgTThrAspValGlnAlaThrValasn 46  
 Db 66 CCTATTGCGCGCATCATGACAAATCTCCACACAGATGTGGAACGACAGTGGT 125  
 QY 47 GlnIleTyAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMet 66  
 Db 126 CAAATTAATCATTTGAACGTGTGGTGCAGATATTGTTCTGCTATCTGTCACAAATG 185  
 QY 67 AspAlaIleGluAlaPheIleTySerGlnGlnValAsnValProIleuValAlaAsp 86  
 Db 186 GATGCGCGGAGCAATTAACAAATTAACAAAGTAAGTTCCTGCTGACAGAT 245  
 QY 87 IleHisPheAspTyrArgIleAlaLeuTyValAlaGluTyrGlyValAspCysLeuArg 106  
 Db 246 ATTCATTTGACATGATGTCGTTAAAGTCGAGATATGAGATGATGTTTACGT 305  
 QY 107 IleAsnProGlyAsnIleGlyAsnGluArgIleArgMetValValAspCysAlaArg 126  
 Db 306 ATCAATCTGCGCAACATGTCGTGAAGATCGGCTCGCGCTGTTGATGTGCGCA 365  
 QY 127 AspIleAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluTyAspLeuGln 146  
 Db 366 GACAAATAATTCGATGCTATTGCTGTAATGACGCGCTTTAAGAAAAAGTTTGCA 425  
 QY 147 GluIleTyGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAsp 166  
 Db 426 GAAATAATATGGCAACCAACGCGCAAGAGCCTGTGTAAATCCCATGCGTATGTAGAA 485

QY 167 HisLeuAspArgLeuAsnPheAspGlnPheValSerValTyAlaSerValaAspValaPhe 186  
 Db 486 ATTCTAATGCTCTTAACCTCGATCAGTTTAAAGTAGAGCTTAAACCTCCGATGATTC 545  
 QY 187 LeuAlaValGluSerTyrArgLeuAlaValGlnIleAspGlnProIleuHisLeuGly 206  
 Db 546 TTAGCGGTGAATCTTATCGTTTACGCTTAAGCAATTAACAGCCTTATCAATTAAGG 605  
 QY 207 IleThrGluAlaGlyGlyAlaArgSerGlyValAlaValysSerAlaIleGlyLeuGlyLeu 226  
 Db 606 ATTCACAGACAGGTGCGACGCGCTGTCGAGTAAATCTGCAGAGGGTTTAAAGATG 665  
 QY 227 LeuIleuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGlu 246  
 Db 666 TTATTACTAGAGGCATTTGGCAGATCACTACGCGTCTTTGGCGGACGATCTGTAGAG 725  
 QY 247 GluIleTyValGlyPheAspIleLeuTySerLeuArgIleArgSerArgGlyIleAsn 266  
 Db 726 GAAATCAAGTCGCTTTGATATTGAAATCTTTACGATTCGTTCAAGAGCAATTAAC 785  
 QY 267 PheIleAlaCysProThrCysSerArgGluPheAspValIleGlyThrValAsnAla 286  
 Db 786 TTATTGCTTGCACCACTGTTCTGCCAGAAATTTGATTAATCGGTACAGTAAATGCG 845  
 QY 287 LeuGlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysVal 306  
 Db 846 CTAGACAAACGCTTGAAGATATATTATACCAATGATGATATATTATTCGTTGTGTA 905  
 QY 307 ValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnTyAla 326  
 Db 906 GTGAATGCTCTGCGAGGCACTGCTCCGATCTCGCGGTACGCGGCGGTAACAAATAA 965  
 QY 327 SerGlyLeuTyArgIleAspGlyValArg--LysAspArgLeuAsnAsnAspMetIle 345  
 Db 966 AGCGGTATTTATCTTACGCGAGCAACGCCAAATAAGCGTTTGTATACGAATATAGTG 1025  
 QY 346 AspGlnLeuGluAlaArgIleArgAlaTyAlaSerGlnLeuAsp 360  
 Db 1026 AACCAATTAGAAGCAAAATTCGTGCGAAATCTCCACGACCAAGAT 1070

# RESULT 9

US-09-252-991A-6420  
 Sequence 6420, Application US/09252991A  
 Patent No. 6551795  
 GENERAL INFORMATION:  
 APPLICANT: Marc J. Rubenfield et al.  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 FILE REFERENCE: 107196.136  
 CURRENT FILING DATE: US/09/252,991A  
 PRIOR FILING DATE: 1999-02-18  
 PRIOR APPLICATION NUMBER: US 60/074,788  
 PRIOR FILING DATE: 1998-02-18  
 PRIOR APPLICATION NUMBER: US 60/094,190  
 NUMBER OF SEQ ID NOS: 33142  
 SEQ ID NO 6420  
 LENGTH: 1644  
 TYPE: DNA  
 ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-6420

Alignment Scores:  
 Pred. No.: 1,586-148 Length: 1644  
 Score: 1372.50 Matches: 267  
 Percent Similarity: 86.46% Conservative: 46  
 Best Local Similarity: 73.76% Mismatches: 48  
 Query Match: 73.55% Indels: 1  
 Gaps: 1

US-09-921-992-78 (1-372) x US-09-252-991A-6420 (1-1644)

QY 1 MethHisAsnGlnAlaProIleGlnArgArgIleTyValGlyAsnVal 20





QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyValAlaArgSerGlyAlaValLysSer 220  
 DB CAGCCCTGACACCTGGGATCACCGAGCGCGCTCGGCTCCGACAGGTGAACCTGC 1468  
 QY 221 AlaIleGlyLeuGlyLeuLeuLeuSerGlyIleGlyAlaSerThrLeuArgValSerLeu 240  
 DB 1467 GCGGTGGGCTGGGACACCTCTCTGCGAGGAAATCGGACACCATTCGGATTCCCTG 1408  
 QY 241 AlaAlaAspProValGluGluIleLeuValAlaGlyPheAspIleLeuLysSerLeuArgIle 260  
 DB 1407 GCTCCGATCGCGTGGAGATCAAGGTGCTTCGATCCTCAAGTCCCTCACCTG 1348  
 QY 261 ArgSerArgIleIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280  
 DB 1347 CGCTCCGCTGGACATCACTTCATCGCTCGCCGAGCTGTGCGGCAAACTTGAGCTG 1288  
 QY 281 IleGlyThrValAsnAlaLeuGluIleArgLeuGluAspIleIleThrProMetAspVal 300  
 DB 1287 GTGAAGACCATGAACGCTGTGAAGGCGCTGGAGACCTCTGTCGATGAGCTG 1228  
 QY 301 SerIleIleGlyCysValValAlaAsnGlyProGlyAlaLeuValSerThrLeuGlyVal 320  
 DB 1227 GCCGTGATCGGTGCGTGGTCAACGCTCGGCGAAGCCAAAGAGCCCATGTGCGCTC 1168  
 QY 321 ThrGlyGlyAsnLysLysSerGlyLeuThrGluAspGlyValArgLysAspArgLeuAsp 340  
 DB 1167 ACCGGCGGCTCCGAGCTG---GTGATATCGACGCGCAAGCCGCTGCAGAAACTGACC 1111  
 QY 341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360  
 DB 1110 AACGACACCTGCTGAGCTGTGAACGCTGATCCGACAGAAAGCGCGGAGAAAGCC 1051  
 QY 361 GluAla 362  
 DB 1050 GAGGCC 1045

# RESULT 11

US-09-252-991A-6339  
 ; Sequence 6339, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 6339  
 ; LENGTH: 2088  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-6339

Alignment Scores:  
 Pred. No.: 2 47e-947 Length: 2088  
 Score: 1363.50 Matches: 266  
 Percent Similarity: 86.59% Conservative: 44  
 Best Local Similarity: 74.30% Mismatches: 47  
 Query Match: 73.07% Indels: 1  
 DB: 4 Gaps: 1

US-09-921-992-78 (1-372) x US-09-252-991A-6339 (1-2088)

QY 5 AlaprolleGlnArgGlySerThrArgIleValGlyAsnValProIleGlyAsp 24  
 DB 5 TCTCGATCATTCGCGGCAAGTCTCGAANAATCTGCTGGCAACGTCCTCCGCTGGCGGC 64

QY 25 GlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAspValGluAlaThr 44  
 DB 65 GAGCGCGCGATTCGCGGTGGAGACATGACCAACCGAGACCTGGCGCTGCGCAC 124  
 QY 45 ValAsnGlnIleLeuAlaLeuGluArgValGlyAlaAspIleValArgValSerValPro 64  
 DB 125 GTCGCGAGATTCGCGCGCTGGAAGATCGCGCGCGGACATCTGCGGGTTCGCTCC 184  
 QY 65 ThrMetAspAlaIleAlaValIlePheLysLeuIleLysGlnGlnValAsnValProLeuVal 84  
 DB 185 GACATGAGACCGCGCGGAGCATTCGCGAATATCAAGCAGAGTCAACGTCGCGCTGTC 244  
 QY 85 AlaAspIleHisPheAspIleArgIleAlaLeuLysValAlaGluThrGlyValAspCys 104  
 DB 245 GCGACATTCACACTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 304  
 QY 105 LeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleThrMetValAlaAspCys 124  
 DB 305 CTGCGCATCAATCGCGGCAACATCGTGGCGAGACCGGTCAAGCGCTGTCGATGCC 364  
 QY 125 AlaArgAspLysAsnIleProIleArgIleGlyValAlaAsnAlaGlySerLeuGluLysAsp 144  
 DB 365 GCGCGGACGCGCAACATTCGCGATTCGCGATTCGCGATTCGCGATTCGCGATTCGCG 424  
 QY 145 LeuGlnGluLysIleGlyLysProThrProGlnAlaLeuLeuGluSerAlaMetArgHis 164  
 DB 425 CTGCAAGAAATATCGCGGCAACCGCGGACCGCTGCGATTCGATTCGATTCGATTCG 484  
 QY 165 ValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAsp 184  
 DB 485 GTGATCATCTCGACACATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 544  
 QY 185 ValPheLeuAlaValGluSerThrArgLeuLeuAlaLysGlnIleAspGlnProLeuHis 204  
 DB 545 GCTTCATGCGCGCTGCGCGCTATCGCTGCGCGCGAGATCGAGCGAGCGCTGCGAC 604  
 QY 205 LeuGlyIleThrGluAlaGlyValAlaArgSerGlyAlaValLysSerAlaIleGlyLeu 224  
 DB 605 CTGGGATCACCGAGCGCGCGCGCTGCGCTGCGCGCGAGATTCGCGCGCTGCGCG 664  
 QY 225 GlyLeuLeuLeuSerGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspPro 244  
 DB 665 GGCATGCTCTCGCGCGCGAGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 724  
 QY 245 ValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGly 264  
 DB 725 GTGAGGAGATCAAGGTGCTTCGACATCTCAAGTCTGACCTGCGCTGCGCTGCG 784  
 QY 265 IleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrVal 284  
 DB 785 ATCAACTTCATCGCTGCGCGCGAGCTGCGCGCGAGAACCTTCGAGTGTGAAGACATG 844  
 QY 285 AsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleGly 304  
 DB 845 AACGAGCTGGAAGGCGCTGCGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 904  
 QY 305 CysValAlaAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyLysAsn 324  
 DB 905 TCGGTGTCACGCTGCGCGGAGACCGAAGAGGCGCATGTCGCGCTACCGCGCGCACT 964  
 QY 325 LysLysSerGlyLeuThrGluAspGlyValArgLysAspArgLeuAspAsnAspMet 344  
 DB 965 CCGAAGCTG---GTGATATGAGCGAGCGAAGCGGTGCGAANAATCTGCGCAACGAC 1021  
 QY 345 IleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAla 362  
 DB 1022 GTGACGAGCTGGAAGCGCTGATTCGCGCAAGAGCGCGCGAGAAAGCGCGAGGCC 1075

# RESULT 12

US-08-827-190-10/C  
 ; Sequence 10, Application US/08827190  
 ; Patent No. 5858367  
 ; GENERAL INFORMATION:



```

APPLICANT: Rather, Philip N.
TITLE OF INVENTION: Methods For Screening For Antimicrobials
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESSES:
ADDRESS: Medlen & Carroll, LLP
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/827,190
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Peter G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: CASE-02443
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 886 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-827-190-10

Alignment Scores:
Pred. No.: 1,11e-146 Length: 886
Score: 1353.00 Matches: 268
Percent Similarity: 96.266 Conservative: 15
Best Local Similarity: 91.164 Mismatches: 11
Query Match: 72.514 Indels: 0
Gaps: 0
DB: 2

US-09-921-992-78 (1-372) x US-08-827-190-10 (1-886)
QY 29 AAlaValGlnSerMetThrAsnThrArgThrAspValGluAlaThrValAsnGlnIle 48
DB 884 GGTGTCATCTATATGAGAAATACGCCACGACGATGTTGAAGCCACTGTGGCGCAATC 825
QY 49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
DB 824 CATACCTTGACGCTGTAGCTGTGATCTGTCGCGTGTCTGTCCTACATGATGATGCA 765
QY 69 AlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHis 88
DB 764 GCAGAACCTTTAAATTAATTAAGACGCGCGTAATGTCCATGTGCGGATATTCAC 705
QY 89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
DB 704 TTTGACTACCGATGCGCATGAAAGTGGCTGATATGCTGATGCTCCTACGATTAAC 645
QY 109 ProGlyAsnIleGlyAsnGluGluArgIleArgMetValAlaAspCysAlaArgAspLys 128
DB 644 CCAGGTAAATATGGAGTGAACAGCGTATTCGCCCAAGCTGTGATGATGCTGATCACC 585
QY 129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnIleLys 148
DB 584 AACATTCCTATCCGATATAGGGCTCATGTGGCGGTCACTGGAGAAAGATATCCAGAAAAA 525
QY 149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
DB 524 TACGGTGAGCCAAACACGTGAACATGTGTTGAATACAGCAATCGACATGTTGATATCTTG 465

```

```

QY 169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
DB 464 GACAGCGTATTCATTCATGATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 405
QY 189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
DB 404 GTCCGCTTATGTTATGTTATGCGCAAAAATGATGATCAACCACTTACCTCGGATATACA 345
QY 209 GluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeu 228
DB 344 GAAGCGGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 285
QY 229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluIle 248
DB 284 GCGAAGAGTATCGGCATACGTTACATCTCTCGCGAGATCTTGTAGGAAGTG 225
QY 249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPhel 268
DB 224 AAAGTCGTTTGAATATCTAAATGCTTACGGATCGCTCAAGTCAACTTAT 165
QY 269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu 288
DB 164 GCTTGCCCACTGTTACGCGCAAGATTTGATGATGATGATGATGATGATGATGATGATG 105
QY 289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn 308
DB 104 CAGCGCTCGAAGATTTTATCCACCGCGAATGATGCTCATTAATGATGATGATGATGATG 45
QY 309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGly 322
DB 44 GCCCGGCGTGAAGCGAGGTTCTTACTTAAAGTGGCGTGC 3

RESULT 13
US-09-170-187-10/C
Sequence 10, Application US/09170187
Patent No. 6383745
GENERAL INFORMATION:
APPLICANT: Rather, Philip N.
TITLE OF INVENTION: Methods For Screening For Antimicrobials
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESSES:
ADDRESS: Medlen & Carroll, LLP
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/170,187
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/827,190
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Peter G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: CASE-02443
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 886 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

```

TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)

US-09-170-187-10

Alignment Scores:

Pred. No.:	1,11e-146	Length:	886
Score:	1353.00	Matches:	268
Percent Similarity:	96.26%	Conservative:	15
Best Local Similarity:	91.16%	Mismatches:	11
Query Match:	72.51%	Indels:	0
		Gaps:	0

US-09-921-992-78 (1-372) x US-09-170-187-10 (1-886)

```

OY 29 AlAValGInserMetThrAsnThrArgThrAspValGluAlaThrValAsnGlnIle 48
DB |||||||
OY 884 GCTGTCCAACTGATGAGCAATACCGCAGCAGATGTGTGAAGCAGCTGTGGCAATTC 825
DB |||||||
OY 49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
DB |||||||
OY 824 CAATCACTGAGCGTGTAGTGTGATATGCTCGCGGTGTGTCTTCCAGATGATGCA 765
DB |||||||
OY 69 AlacIuaIaPheIysLeuIleLysGlnIleValAsnValProLeuValAlaAspIleHis 88
DB |||||||
OY 764 GCAGAGCGCTTAAATTAATTAACAGCGCGTGAATGCGCATTTGGTGGATATTCTAC 705
DB |||||||
OY 89 PheAspTYRArgIleLeuIleValAlaGluIuTYRValAlaAspCysLeuArgIleAsn 108
DB |||||||
OY 704 TTTGACTACCTGATCGGATGATAAGTGGCTGAATATGCTGTGACTGCTTACGATATAC 645
DB |||||||
OY 109 ProGluAsnIleGlyAsnGluIuArgIleArgMetValValAspCysAlaArgAspIlys 128
DB |||||||
OY 644 CCAGGTATATCGGAGCGAGAGCGATTCGCCAAGTCTGTGATAGTCTGTCATAC 585
DB |||||||
OY 129 AsnIleProIleArgIleGlyValAlaAsnIleGlySerLeuGluIuLysAspLeuGlnIuLys 148
DB |||||||
OY 564 AACATTCCTATCCGATTAAGGGGTCAATGGCGGTCACTGCGAATAATATCCAGAAAAA 525
DB |||||||
OY 149 TYRGIuIuProIleProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
DB |||||||
OY 524 TACGGTAGCCAAACCGTGAAGCATTTGGTGAATGCAGCAATGCCAGATGTGATATCTTG 465
DB |||||||
OY 169 AspArgIleAsnPheAspGlnPheIysValSerValLysAlaSerAspValPheLeuAla 188
DB |||||||
OY 464 GACAGGCGTAATTCGATCAGTCAAGGTCAAGTGTAAAGCGTGGAGATGCTTTCTTCC 405
DB |||||||
OY 189 ValGluSerTYRArgIleLeuAlaLysGlnIleAspGlnProLeuHisLeuGluIleThr 208
DB |||||||
OY 404 GTGGCGCTTATTCGTTATGCGCGCAAAAATATGATCAACCACTCCACCTGGTATTACA 345
DB |||||||
OY 209 GluAlaGlyAlaArgSerGlyValAlaLysSerAlaIleGlyLeuGluLeuLeu 228
DB |||||||
OY 344 GAAAGCGGTGGCGCTGTCTGTTCAAGTGAATCAGCAATTTGGTCTGTGATTTGTTGG 285
DB |||||||
OY 229 SerGIuGIuIleGlyAspThrLeuArgValSerLeuAlaIleAspProValGluIuLe 248
DB |||||||
OY 284 GGTGAGGATTCGGGATCAGTTACGATCTCAGTGGCGGAGATCCTGTTGAGGAAGTG 225
DB |||||||
OY 249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
DB |||||||
OY 224 AAAGTCGTTTGAATATCTTAAATCGTTACGATCCGCTCACGGGATCAACTTATAT 165
DB |||||||
OY 269 AlaCysProIleCysSerArgGlnIuPheAspValIleGlyThrValAlaAsnLeuGlu 288
DB |||||||
OY 164 GCTTGCCCAACTGTGTGCGCCAAAGATTTGATGTGATGTGATACGTAATCTTTGGAG 105
DB |||||||
OY 288 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValAlaAsn 308
DB |||||||
OY 104 CAGCCCTTCAGAGATATATCAGCCGAGATGATGCTCATTAATTTGGTGTGTAGAGAT 45
DB |||||||
OY 309 GlyProGluIuAlaLeuValSerThrLeuGlyValThrGly 322
DB |||||||
OY 44 GGCCTGGGATGAAGCCGAGGTTTCTACTTAAAGTGTGGCTGGC 3
DB |||||||

```

# RESULT 14

US-09-328-352-3780

; Sequence 3780, Application US/09328352

; Patent No. 6562938

; GENERAL INFORMATION:

; APPLICANT: Gary L. Bretton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; TITLE OF INVENTION: BAUMANNIT FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328,352

; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 3780

; LENGTH: 1137

; TYPE: DNA

; ORGANISM: Acinetobacter baumannii

US-09-328-352-3780

Alignment Scores:

Pred. No.:	2.67e-128	Length:	1137
Score:	1195.00 <td>Matches:</td> <td>231 </td>	Matches:	231
Percent Similarity:	81.77% <td>Conservative:</td> <td>65 </td>	Conservative:	65
Best Local Similarity:	63.81% <td>Mismatches:</td> <td>66 </td>	Mismatches:	66
Query Match:	64.04% <td>Indels:</td> <td>0 </td>	Indels:	0
		Gaps:	0

US-09-921-992-78 (1-372) x US-09-328-352-3780 (1-1137)

```

OY 6 ProIleGlnArgArgIysSerThrArgIleTYRValGlyAsnValProIleGlyAspGly 25
DB |||||||
OY 34 CCAATTAAAGCTCCACCAACACGTAATTCGTTGTGTCGCTATGTCGGTGGCAT 93
DB |||||||
OY 26 AlaProIleAlaValGlnSerMetThrAsnThrArgThrAspValGluAlaThrVal 45
DB |||||||
OY 94 GCACCTTATAGTGCAGAAAGTATGACAAATCCCAACTTGCATGTGGACGCAACTGTG 153
DB |||||||
OY 46 AsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThr 65
DB |||||||
OY 154 GCTAGATGAGCGTGTGTGTTGATGACAGGTGCAGATATATAGCTGTTCAGTCCCTCT 213
DB |||||||
OY 66 MetAspAlaIleAlaValPheLysLeuIleLysGlnIleValAsnValProLeuValAla 85
DB |||||||
OY 214 ATGAGAGCTGCTGAAGCCTTTGGTGCATCCGTAAGCGTGTTCAGTTCCATTTAGTACT 273
DB |||||||
OY 86 AspIleHisPheAspTYRArgIleAlaLeuLysValAlaGluTYRGIuValAspCysLeu 105
DB |||||||
OY 274 GATATCCATTTTGAACCAATGAATGCTTGGCAGTTCGAGATATAGGTGGCGACTGCTTG 333
DB |||||||
OY 106 ArgIleAsnProGluAsnIleGlyAsnGluIuArgIleArgMetValAlaAspCysAla 125
DB |||||||
OY 334 CGTATTAAACCCGGGCAATATCGTTCAGACAGCAAAAGTTGCTGAAGTTGGTGGCGGCA 393
DB |||||||
OY 126 ArgAspIysAsnIleProIleArgIleGlyValAlaAsnIleGlySerLeuGluLysAspLeu 145
DB |||||||
OY 394 CGTATCAACGGTATTTCTATGCGTATGCGGTGATGGCGGTCTCTAGAAAAAGATTTA 453
DB |||||||
OY 146 GlnGluIuTYRGIuProIleProIleAlaLeuLeuGluSerAlaMetArgHisVal 165
DB |||||||
OY 454 CAGAAAAAATATGGGACCTTACAGGGCAGGACACTCTTGAGTCACTTTACGTCATATT 513
DB |||||||
OY 166 AspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspVal 185
DB |||||||
OY 514 GATATTATTAAGCCGCTTGACTTCATGAGTTTAAAGTCACTGTAAAGATCAATATGTG 573
DB |||||||
OY 186 PheLeuAlaValGluSerTYRArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeu 205
DB |||||||
OY 574 TTTTAAACCATGAGATCTTATCGTTACTCTCTCAACAAATGTGATTAACCATTCACCTT 633
DB |||||||
OY 206 GlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGly 225
DB |||||||
OY 634 GGAGTACTGAAGCTGATTTTACCGTACAGGTACTGTGAATAGCGGATTTGCTTTGGT 693
DB |||||||

```

QY 226 LeuLeuLeuSerGluGluGlyIleGlyAspThrLeuArGValSerLeuAlaIleAspProVal 245  
 Db 694 GGGTATATGATGAAAGCATTTGGCGATACGATGGCTATTCGCTGCTGCTGAACTGAA 753  
 QY 246 GluGluIleLeuValGlyPheAspIleLeuLeuSerLeuArGlyIleArGlyIle 265  
 Db 754 GATGAATCAAGACGCTTTGATATCTTAATTAACGCTTGCGCTTCAACGGATC 813  
 QY 266 AsnPheIleLeuValProThrCysSerArGValGluPheAspValIleGlyThrValAsn 285  
 Db 814 AACTTATGCTTCTCCAAAGTGTCTTCCGCAAAATTAACGATTCAGTGGATGAC 873  
 QY 286 AlaLeuGluArGValGluAspIleIleThrProMetAspValSerIleIleGlyCys 305  
 Db 874 GCTTAAAGAGCGCTTGAAGATATCCGATACACCAATGACGCTTGGCTATGCTGT 933  
 QY 306 ValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyIleAsn 325  
 Db 934 AAGGTAAATGCGCCAGGTGAAGCAAAAGACAGATATCGGGTGTGGCGCTCGCC 993  
 QY 326 LysSerGlyLeuArgIleAspGlyValArgIleAspArgLeuAspAsnAspMetIle 345  
 Db 994 CGTTCATGTTGTTATCTATGATGATGAAGAAAGCCATTATATCAATCAATGATG 1053  
 QY 346 AspGluLeuGluAlaArgIleArgIleAlaValSerGluLeuAspGluAlaArgArgIle 365  
 Db 1054 GATGAATCAAGACCTTGTGTTCAACGTTCAACGCTTGAAGACCTTAATATATA 1113  
 QY 366 AspVal 367  
 Db 1114 GAAAT 1119  
 RESULT 15  
 US-08-827-190-9/c  
 Sequence 9, Application US/08827190  
 Patent No. 5858367  
 GENERAL INFORMATION:  
 APPLICANT: Rather, Philip N.  
 TITLE OF INVENTION: Methods For Screening For Antimicrobials  
 TITLE OF INVENTION: Utilizing aarc And Compositions Thereof  
 NUMBER OF SEQUENCES: 13  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Medlen & Carroll, LLP  
 STREET: 220 Montgomery Street, Suite 2200  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: United States of America  
 ZIP: 94104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/827,190  
 FILING DATE:  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Carroll, Peter G.  
 REGISTRATION NUMBER: 32,837  
 REFERENCE/DOCKET NUMBER: CASE-02443  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 705-8410  
 TELEFAX: (415) 397-8338  
 INFORMATION FOR SEQ ID NO: 9:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 866 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: unknown  
 MOLECULE TYPE: DNA (genomic)  
 US-08-827-190-9

Alignment Scores:  
 Pred. No.: 2,68e-84 Length: 886  
 Score: 812.00 Matches: 158  
 Percent Similarity: 70.65% Mismatches: 49  
 Best Local Similarity: 53.92% Indels: 0  
 Query Match: 43.52% Gaps: 0

US-09-921-992-78 (1-372) x US-08-827-190-9 (1-886)

QY 30 ValGlnSerMetThrAsnThrArgThrAspValGluAlaThrValAsnGlnIleLeu 49  
 Db 881 ATCCAAAGCATGACAAACAAACAAACACATGATGAGAAGCAACGCTGCGGAATTAAC 822  
 QY 50 AlaLeuGluArGValGlyAlaAspIleValArgValSerValProThrMetAspAla 69  
 Db 821 CGTTGGCTGAGAGCGCGATGCCAAATTCCTGCGGATGACATGCGGATGAACGGCGCA 762  
 QY 70 GluAlaPheLeuIleLeuGlnGlnValAsnValProLeuValAlaAspIleIlePhe 89  
 Db 761 AAGCCATTGCGGATATTAAGCAAGCATTTCCATTCCTGCTGCTGATGACATTCAT 702  
 QY 90 AspTyrArgIleAlaLeuValAlaGluIleValArgValAspCysLeuArgIleAsnPro 109  
 Db 701 GATTAATAACTGCGCTGTAAGCCATTGAAGCGCGAGATTAATAATCCGATCAACCC 642  
 QY 110 GlysAnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLysAsn 129  
 Db 641 GGCATATTCGCGCGCGCGCAAAAGTTGAACGCTGTTAAAGCGCAAAAGCAAGCC 582  
 QY 130 IleProIleArgIleGlyValAlaAsnIleGlySerLeuGluLysAspLeuGlnLysTyr 149  
 Db 581 ATCCGATCAGAAATCGGATTAACCGCGCTTCATTTGAAACGGAATTTAGAAAATAC 522  
 QY 150 GlyLupProThrProGlnAlaLeuLeuGlnSerAlaMetArgHisValAspHisLeuAsp 169  
 Db 521 GGTATCCGACTGCCGATGATGATGATGATGATGATGATGATGATGATGATGATG 462  
 QY 170 ArgLeuAsnPheAspGlnPheValSerValLysAlaSerAspValPheLeuAlaVal 189  
 Db 461 GATCTTGAATTTTCACGATATTTATGTCAGTGAAGCCCTGTGACGTAACCTTGCA 402  
 QY 190 GluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThrGlu 209  
 Db 401 GAGGCTTATGAAGAAAGACGAGGAAGCGTTGACTACCCCTTACCTCGGATCACCGAG 342  
 QY 210 AlaGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuSer 229  
 Db 341 TCAGGAACACTGTTGGCGCGCACAGTAAAGAGCGACAGCACTCGCGCATTTTAAGC 282  
 QY 230 GluGlyIleGlyAspThrLeuArgValSerLeuAlaAspProValGluGluIleLeu 249  
 Db 281 AAAGCATCGGGAACACATGCGCATTTCTAAGCCACACACCTGAGGAAGGTTAA 222  
 QY 250 ValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgIleAsnPheIleAla 269  
 Db 221 GTAGCAAGAGAGCTTGAATCTTTCGCTTACCTCCAAATCGCGACGCACTCA 162  
 QY 270 CysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGluGln 289  
 Db 161 TGCCGCGCTGCGCGCTATGATGATGATGATGATGATGATGATGATGATGATGATG 102  
 QY 290 ArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsnGly 309  
 Db 101 TATATTTCTAAGATAAAGCGCGATTAAGTATGATGATGATGATGATGATGATGATG 42  
 QY 310 ProGlyGluAlaLeuValSerThrLeuGlyValThrGly 322  
 Db 41 CTTGATGAGCGAGAGACGCTGATATCGCAATCGCGCGC 3

Search completed: August 15, 2003, 02:02:49  
 Job time: 1575 secs

Fri Aug 15 14:34:49 2003

us-09-921-992-78.rni

Page 14

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OK protein - nucleic search, using frame\_plus\_p2n model

Run on: August 15, 2003, 01:35:19 ; Search time 326 Seconds  
(without alignments)  
2553.723 Million cell updates/sec

Title: US-09-921-992-78

Perfect score: 1866

Sequence: 1 MHNQAPIQRKSTRIVGVN.....RAKASQLDEARRIDVQVEX 372

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1504479 seqs, 1118970152 residues

Total number of hits satisfying chosen parameters: 3008958

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=slh  
-O/cg2\_1/USPTO.spool/US09921992/runtat\_12082003.094917.17547/app\_query.fasta\_1.519  
-DB=published.Applications\_NA -QFMT=fastcap -SUFFIX=rmph -MINMATCH=0.1  
-LOOPCT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.ccl -LIST=45 -DOCALIGN=200 -THR.SCORF=pct -THR.MAX=100  
-THR.MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEATSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09921992.ecgn\_1.1.233-@runtat\_12082003.094917.17547  
-NCPU=6 -ICPU=3 -NO.MMP -LARGEQUERY -NEG.SCORF=0 -WAIT -DSPLOCK=100  
-LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications\_NA:\*

1: /cg2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cg2\_6/ptodata/1/pubpna/PTI\_NEW\_PUB.seq:\*  
3: /cg2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cg2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cg2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cg2\_6/ptodata/1/pubpna/PTUS\_PUBCOMB.seq:\*  
7: /cg2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cg2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cg2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cg2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cg2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cg2\_6/ptodata/1/pubpna/US09C\_NEW\_PUB.seq:\*  
13: /cg2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
14: /cg2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cg2\_6/ptodata/1/pubpna/US10B\_NEW\_PUB.seq:\*  
16: /cg2\_6/ptodata/1/pubpna/US10C\_NEW\_PUB.seq:\*  
17: /cg2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1866	100.0	1119	9	US-09-921-992-3 Sequence 3, Appl1

Result No.	Score	Query Match	Length	DB ID	Description
2	1603.5	85.9	1830121	14	US-10-329-960-1 Sequence 1, Appl1
3	1282	68.7	640681	10	US-09-790-988-1 Sequence 1, Appl1
4	822	44.1	1083	10	US-09-974-300-1692 Sequence 1692, Ap
5	808	43.3	1155	14	US-10-156-761-2548 Sequence 2548, Ap
6	808	43.3	9025608	14	US-10-156-761-1 Sequence 1, Appl1
7	794	42.6	1180	10	US-09-881-7524-195 Sequence 195, Ap
8	786.5	42.1	1155	14	US-10-156-761-1636 Sequence 1636, Ap
9	761	40.8	1134	10	US-09-738-626-2205 Sequence 2205, Ap
10	751.5	40.3	3309400	10	US-09-738-626-110 Sequence 110, Ap
11	462	24.8	2520	14	US-10-066-543-1184 Sequence 1, Appl1
12	452	24.2	706	14	US-09-921-992-22 Sequence 22, Appl
13	339	18.2	670	9	US-09-921-992-23 Sequence 23, Appl
14	326	17.5	596	9	US-09-921-992-21 Sequence 21, Appl
15	321	17.2	584	9	US-09-921-992-21 Sequence 21, Appl
16	321	15.6	33675	9	US-09-921-992-21 Sequence 21, Appl
17	277	14.8	705	9	US-09-921-992-34 Sequence 34, Appl
18	277	14.0	601	9	US-09-921-992-11 Sequence 11, Appl
19	262	14.0	528	9	US-09-921-992-15 Sequence 15, Appl
20	259	13.9	528	9	US-09-921-992-16 Sequence 16, Appl
21	258	13.8	379	9	US-09-921-992-36 Sequence 36, Appl
22	243	13.0	511	9	US-09-921-992-36 Sequence 36, Appl
23	241	12.9	498	9	US-09-921-992-24 Sequence 24, Appl
24	230	12.3	403	9	US-09-921-992-24 Sequence 24, Appl
25	221	11.8	938	9	US-09-921-992-13 Sequence 13, Appl
26	218	11.7	395	9	US-09-921-992-13 Sequence 13, Appl
27	215.5	11.5	464	9	US-09-921-992-33 Sequence 33, Appl
28	214	11.4	432	9	US-09-921-992-17 Sequence 17, Appl
29	212	11.4	395	9	US-09-921-992-35 Sequence 35, Appl
30	210	11.3	564	9	US-09-921-992-35 Sequence 35, Appl
31	198.5	10.6	443	9	US-09-921-992-42 Sequence 42, Appl
32	192.5	10.3	293	9	US-09-921-992-42 Sequence 42, Appl
33	179	9.6	388	9	US-09-921-992-45 Sequence 45, Appl
34	174	9.3	440	9	US-09-921-992-48 Sequence 48, Appl
35	170.5	9.1	211	9	US-09-921-992-46 Sequence 46, Appl
36	166.5	8.9	436	9	US-09-921-992-46 Sequence 46, Appl
37	166	8.9	869	10	US-09-974-300-6102 Sequence 6102, Ap
38	157.5	8.4	368	9	US-09-921-992-42 Sequence 42, Appl
39	155.5	8.3	551	9	US-09-921-992-41 Sequence 41, Appl
40	155	8.3	430	9	US-09-921-992-49 Sequence 49, Appl
41	155	8.3	615	9	US-09-921-992-6 Sequence 6, Appl
42	154.5	8.3	619	9	US-09-921-992-27 Sequence 27, Appl
43	151	8.1	422	9	US-09-921-992-28 Sequence 28, Appl
44	150	8.0	589	9	US-09-921-992-7 Sequence 7, Appl1
45	149.5	8.0	399	9	US-09-921-992-47 Sequence 47, Appl

## ALIGNMENTS

RESULT 1  
US-09-921-992-3  
Sequence 3, Application US/09921992  
Patent No. US20020069426A1  
GENERAL INFORMATION:  
APPLICANT: Boronati, Albert;  
APPLICANT: Campos, Narciso;  
APPLICANT: Rodriguez-Concepcion, Manuel;  
APPLICANT: Rohmer, Michel;  
APPLICANT: Seeman, Myriam;  
APPLICANT: Valentini, Henry E.;  
APPLICANT: Venkatesh, Tyamagondlu V.;  
TITLE OR INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes  
FILE REFERENCE: 16516.107/35-21(51897)US  
CURRENT APPLICATION NUMBER: US/09/921,992  
PRIOR APPLICATION NUMBER: 2001-08-06  
PRIOR FILING DATE: 2000-08-07  
NUMBER OF SEQ ID NOS: 85  
SEQ ID NO 3  
LENGTH: 1119  
TYPE: DNA  
ORGANISM: Escherichia coli  
FEATURE:

NAME/KEY: CDS  
LOCATION: (1) (1119)  
US-09-921-992-3

Alignment Scores:  
Pred. No.: 3,17e-216 Length: 1119  
Score: 1866.00 Matches: 372  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0

US-09-921-992-78 (1-372) x US-09-921-992-3 (1-1119)

```

OY 1 MethisnnglnalaproileglnargarylseserThrarglyValasrval 20
DB 1 ATGCATACCGAGGCTCCATCAATCAAGAAATCAACAGCTATTTCAGTGGGAATGTG 60
OY 21 ProileglyaspiglyalaproilealavalglnsermethrasnthrarqThrThrasp 40
DB 61 CCGATTGGCGATGGTGTCCCATCGCTACAGTCACATGACCAATACGCTAGCAGAC 120
OY 41 Valglualarhrvalasnlnlelysalaleuglnuargvalaglyalaspillevalarg 60
DB 121 GTCGAGACGAGGTCATCAATCAATCAAGCGCTGGAGCGCTTGCGCTGATATCGTCGT 180
OY 61 ValserValProthmetaspAlaaglyalalelylsleuileysglnlnvalasp 80
DB 181 GTATCCGTACCGACGATGAGCGCGCAAGCGTTCAACATCAACACAGAGTTAC 240
OY 81 ValproleuValAlaspillehsphesapTyarargylealeuysValalaglulTy 100
DB 241 GTCCCGCTGGTGGCGGATCATTCGACATCGCATTCGCGTAAAGTACGGAATAC 300
OY 101 GlyalaspCyaleuargileasnProglyasnilleglyasnuglnuargileargmet 120
DB 301 GCGCGCATGTGCTGCTGATTAAACCTGCAATATCGGTAAAGACGATTCGCATG 360
OY 121 ValValaspCyAlaargAspIysasnilleProileargyleValasrvalaglyser 140
DB 361 GTGGTGGATGTGGCGCGGATTAACATTCGATCCGATTGGCGTTAACGCGGATCG 420
OY 141 IeuglulysaspIeuglnlulysTyrglylupProthProglAlaleuIeuglnser 160
DB 421 CTGGAAGAAAGATCTCAAGAAAGATGCGAGAACCGCGCGCGCTGCTGGAATCT 480
OY 161 AlametarghlsValasphlsleuaspargleuasnphesapglPhelysValserVal 180
DB 481 GCCATGCGCATGTGATCATCTCGATCCCTGGAACCTTCGATCAAGTCAAGCGTG 540
OY 181 LysAlaseraspValphelenuAlavalglnserTyarargleuenuAlalysglnleasp 200
DB 541 AAGCGCTGACGCTCTCTCGCTGTAGTCTTATCGTTCTGCAAAACAGATCGAT 600
OY 201 GlnProleuHlsleuglylethrGlnaglyAlaargSerGlyAlaVallysser 220
DB 601 CAGCCGTTCACTGGGATCACCGAGCGGTGGCGGAGGCGGAGTAAATCC 660
OY 221 AlaileglyleuIeuleuIeuserGlnlyleglYaspThrleuargValserleu 240
DB 661 GCCATGCTTACGTGCTGCTGCTGCTGAGGATGGGACACGCTGGCGATCGCTG 720
OY 241 AlaAlaaspProvalglnulileysValglyPhaspilleuIeulysSerleuargyle 260
DB 721 GCGCGCGATCGGTCGAAGATCAAGATCGGTTTCGATATTTTGAATTCGCGCTATC 780
OY 261 ArgSerArglyleasnphellealacysProthCysSerArgglnlnpheaspVal 280
DB 781 CGTTCCGAGGATCACTTCATCGCTGCGCACTGTTCCGCTAGGAATTTGATGTT 840
OY 281 IleglyThrValasnAlaleuglnuargleuglnuaspilleleThrPrometaspVal 300
DB 841 ATGCGTACGTTAACCGCGCTGAGCAACGCTCGAAGATATCATCTCGATGAGCGTT 900

```

```

OY 301 serlleleGlyCysValValasnGlyProglyGlnAlaleuValserThrleuglyVal 320
DB 901 TCGATTATCGCTGCTGCTGATGATGCCAGGTGAGCGCGCTGTTCTACATCGCGCTC 960
OY 321 ThrGlyGlyAsnIysSerGlyleuTyrglyuaspGlyValalarglyaspProleuasp 340
DB 961 ACCGCGGCAACAAAGAAAGAGCGCCCTCTATGAAGAGGCGTCCGCAAGACCGCTGAGC 1020
OY 341 AsnAspAspmetlleaspGlnleuglnuargileargAlalysAlaserGlnleuasp 360
DB 1021 AACACGATATGATGACACGCTGGAACGACGATTCGCGCAAGCCAGTCAGCTGAGC 1080
OY 361 GlualaargyleaspValaspValglnlnValglnlys 372
DB 1081 GAAGCGCTGCAATTCGCTTCAGCAGGTTGAAGAA 1116

```

# RESULT 2

US-10-329-960-1

Sequence 1, Application US/10329960  
Publication No. US20030099277A1

GENERAL INFORMATION:

APPLICANT: Fleischmann et al.

TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, F

FILE REFERENCE: PB186PI

CURRENT APPLICATION NUMBER: US/10/329,960

PRIOR APPLICATION NUMBER: US 08/643,990

PRIOR FILING DATE: 2000-08-23

PRIOR APPLICATION NUMBER: US 08/487,429

PRIOR FILING DATE: 1995-06-07

PRIOR APPLICATION NUMBER: US 08/426,787

PRIOR FILING DATE: 1995-04-21

NUMBER OF SEQ. ID NOS: 1

SOFTWARE: PatentIn version 3.1

SEQ. ID NO. 1

LENGTH: 1830121

TYPE: DNA

ORGANISM: Haemophilus influenzae

FEATURE:

NAME/KEY: misc\_feature

LOCATION: (4747)..(4747)

OTHER INFORMATION: n equals a, t, g or c

FEATURE:

NAME/KEY: misc\_feature

LOCATION: (9921)..(9921)

OTHER INFORMATION: n equals a, t, g or c

FEATURE:

NAME/KEY: misc\_feature

LOCATION: (10150)..(10150)

OTHER INFORMATION: n equals a, t, g or c

FEATURE:

NAME/KEY: misc\_feature

LOCATION: (29298)..(29298)

OTHER INFORMATION: n equals a, t, g or c

FEATURE:

NAME/KEY: misc\_feature

LOCATION: (36543)..(36543)

OTHER INFORMATION: n equals a, t, g or c

FEATURE:

NAME/KEY: misc\_feature

LOCATION: (36551)..(36551)

OTHER INFORMATION: n equals a, t, g or c

FEATURE:

NAME/KEY: misc\_feature

LOCATION: (36636)..(36636)

OTHER INFORMATION: n equals a, t, g or c

FEATURE:

NAME/KEY: misc\_feature

LOCATION: (40808)..(40810)

OTHER INFORMATION: n equals a, t, g or c

```
NAME/KEY: misc_feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (53569)..(53569)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152530)..(152530)
```

Alignment Scores: 1.39e-179 Length: 1830121  
 Pred. No.: 1603.50 Matches: 320  
 Score: 93.66% Conservative: 20  
 Percent Similarity: 93.66% Mismatches: 22  
 Best Local Similarity: 88.15% Indels: 1  
 Query Match: 85.93% Gaps: 1  
 DB: 14

US-09-921-992-78 (1-372) x US-10-329-960-1 (1-1830121)

QY 4 GlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnValProIleGly 23  
 DB 390972 CAGCCAACTATTGAAGCGTCGTAATCGCAAAAATTTATGGGAAATGTACCATGGT 391031  
 QY 24 AspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAspValGluAla 43  
 DB 391032 GGGATGCGCTATTATTCGGCTCATTCATGCAAAATCTCGCACCATGATGTGAAGCG 391091  
 QY 44 ThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerVal 63  
 DB 391092 ACAGTTCCTCAAAATTAATTCATTGAACGCTGTGCGAGATATGTTCGTATCTGTT 391151  
 QY 64 ProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeu 83  
 DB 391152 CCACAAATGATGCTGCGGAGCATTTAAACAATTAACAACAAGTGAATGTTCCGCTC 391211  
 QY 84 ValAlaAspIleLysPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAsp 103  
 DB 391212 TAGACAGATATTCATTCGACATTCGATTCGCTTAAAGTCGCAAGATATGAGATGAT 391271  
 QY 104 CysLeuAlaGlyIleAsnProGlyAsnIleGlyAsnGluArgIleArgMetValValAsp 123  
 DB 391272 TGTATTACATCAATTCGCGCAATGCTGCTGAGATGCGGTCGCTGCTGCTGCTGCT 391331  
 QY 124 CysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGlyLys 143  
 DB 391332 TGTGGCGGAGCAAAAATATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 391391  
 QY 144 AspLeuGlnGluLysTyrGlyLysProThrProGlnAlaLeuGlnLysSerAlaMetArg 163  
 DB 391392 GATTTGCAAGAAATATGCGCACCCACCGCAAGAGCCTTTTGAATCCCATGCTGCT 391451  
 QY 164 HisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSer 183  
 DB 391452 CATGTGCAATTCGATTCGCTTAACTTCGATTCGATTCGATTCGATTCGATTCGATTC 391511  
 QY 184 AspValPheLeuAlaValGlnSerTyrArgLeuLeuAlaLysGlnLysPheProLeu 203  
 DB 391512 GATGTATCTTCGCGGCTTAATCTTATGCTTAAAGCAATTAACAAGCCTTTA 391571  
 QY 204 HisLeuGlyIleThrGlnAlaGlyAlaArgSerGlyAlaValLysSerAlaIleGly 223  
 DB 391572 CATTTAGCATTCACACAGAGTGGCGACGCGCTGCGAGTAATAATCTCGAGTGGCT 391631  
 QY 224 LeuGlyLeuLeuLeuSerGlnGlyIleGlyAspThrLeuArgValSerLeuAlaAsp 243  
 DB 391632 TTAGGATGTTTATCTGCTGCGGCTTGGCGATTCACCTACGCTCTCTTGGGCGAGAT 391691  
 QY 244 ProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArg 263  
 DB 391692 CCGTGTAGAGAAATCAAAAGCGGTTTGAATTTTGAATCTTTACGGATTCCTTAAAGA 391751  
 QY 264 GlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThr 283  
 DB 391752 GGAATTAATCTTATGCTTCCCAACCTGTCGCGCAAGAAATTTGTAATCGGTACCA 391811  
 QY 284 ValAsnAlaLeuGlnGluArgLeuGluAspIleIleThrProMetAspValSerIleIle 303  
 DB 391812 GTAATGCGCTAGAACCAACCCCTTGAAGATATTAACCAATGATGATATATATATC 391871  
 QY 304 GlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValIleThrGlyGly 323  
 DB 391872 GATTGTGTAGTGAATGCTTCGCGGAGCACTGCTCTCCGATTCGCGCTAACAGGCGGT 391931

DB 391872 GATTGTGTAGTGAATGCTTCGCGGAGCACTGCTCTCCGATTCGCGCTAACAGGCGGT 391931  
 QY 324 AsnLysLysSerGlyLeuTyrGluAspGlyValArg--LysAspArgLeuAspAsn 342  
 DB 391932 AACCAAAAAAGCGGCTTATTAATCTTACGAGAGAACCGCAAAAAGCGGTTTATTAACGAA 391991  
 QY 343 AspMetIleAspIleGlnGluAlaValGlyIleArgValAlaSerGlnLeuAspGluAla 362  
 DB 391992 GATATGTGTGAACCAATTAAGCAAAAATTCGTCGCAAGGTGCGACAGCAAGTCCAAA 392051

QY 363 ArgArgIle 365  
 DB 392052 AACCAAAAT 392060

RESULT 3

US-09-790-988-1  
 : Sequence 1, Application US/09790988  
 : Patent No. US20020127687A1  
 : GENERAL INFORMATION:  
 : APPLICANT: SHIGENOBU, SHUJI  
 : APPLICANT: MATANABE, HIDEMI  
 : APPLICANT: HATTORI, MASAHIRO  
 : APPLICANT: SAKAKI, YOSHIYUKI  
 : TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS  
 : FILE REFERENCE: 081356/0159  
 : CURRENT APPLICATION NUMBER: US/09/790, 988  
 : PRIOR FILING DATE: 2001-02-23  
 : PRIOR APPLICATION NUMBER: JP2000-107160  
 : NUMBER OF SEQ ID NOS: 7  
 : SOFTWARE: PatentIn Ver. 2.1  
 : SEQ ID NO 1  
 : LENGTH: 640681  
 : TYPE: DNA  
 : ORGANISM: Buchnera sp.  
 : US-09-790-988-1

Alignment Scores: 2.69e-141 Length: 640681  
 Pred. No.: 1282.00 Matches: 242  
 Score: 83.88% Conservative: 65  
 Percent Similarity: 66.12% Mismatches: 57  
 Best Local Similarity: 68.70% Indels: 2  
 Query Match: 10 Gaps: 2  
 DB: 10

US-09-921-992-78 (1-372) x US-09-790-988-1 (1-640681)

QY 1 MethAsnGlnAlaPro--IleGlnArgArgLysSerThrArgIleTyrValGlyAsn 19  
 DB 314272 ATGATTAATTAAGTGTAAATTAATCAATGAGAAATCTGATTCATTATGTTGGAATA 314331  
 QY 20 ValProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThr 39  
 DB 314332 GTGCGTATTGGCAAAATGATGCGCAATATCAGTTCAATCTATGACAAATACGTAATCT 314391  
 QY 40 AspValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleVal 59  
 DB 314392 AATATCTCGAACAATTAATCAAAATCTTACAGTTACCAAAAAGTGAATGATGATGTT 314451  
 QY 60 ArgValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnVal 79  
 DB 314452 CGTATTTCTATCCCAATTTAAAGCTGCGAGATATTCGAAGAAATTAAGAAACAAACA 314511  
 QY 80 AsnValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGlu 99  
 DB 314512 AATGTTCATTAATGACATTAATTAATTTGATTCAGATTCGTTTCAACGCTTAAATA 314571  
 QY 100 TyrGlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGlnGluArgIleArg 119  
 DB 314572 TATGCGAGCATGTTTAAAGATTAATCTCGGGAATATGGAATTAAGAAAGAGATATCA 314631  
 QY 120 MetValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGly 139  
 DB 314631



```

Db      314632 GAATCATTTCTTACGCCAAGATGAAATATTCATTCGATTCGTTAATGCTGGA 314691
QY      140 SerLeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuGln 159
Db      314692 TCTTTAGAAAAAGATATATTAATAAATATTAATACCTACTCCGATGCAATGTGAGAA 314751
QY      160 SerAlaMetLysGlnValAspHisLeuSparGlyLeuAsnPhaeSparGlnPheLysValSer 179
Db      314752 TCAGCTATGAGCAATATGGAATACTGTGATGCTTAAATTTAATCAATTTAAAGTTC 314811
QY      180 ValLysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLeuGln 199
Db      314812 GTTAAAGCGCTGATGATTTTATTTAGCTATTAATCAATTCATTCGATTTAGAAAAAAT 314871
QY      200 AspGlnProLeuHisLeuGlyLLeThrGluAlaGlyAlaArgSerGlyValAlaLys 219
Db      314872 ACACAACTTTCATATATGCGATTAACGATCCGTTGGTTAAGCAATGCAACGTTAAA 314931
QY      220 SerAlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSer 239
Db      314932 TCATCTATGATATGCTTATTTATTTATTAAGAGCATTTGAGATACAAATACGAGTTCA 314991
QY      240 LeuAlaAlaAspProValGluGluGluLysValGlyPheAspIleLeuLysSerLeuArg 259
Db      314992 TTACGGGCACTCCCACTCAAGAAAGTAAAGTACGTTATGCAATTTTAAAGTTTATCT 315051
QY      260 IleArgSerArgGlyIleAsnPhelAlaCysProThrCysSerArgGlnGluPheAsp 279
Db      315052 TTAAGACCAAGAGTAAATTTATTTATTTATGCTGCTCCTGCTTACACAAAGATTGAT 315111
QY      280 ValIleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAsp 299
Db      315112 GTATTTAATACAGTAATCAACATGAAATAAATCTAAGATATCTCCATCCCATGAT 315171
QY      300 ValSerIleIleGlyCysValAlaAsnGlyProGlyGluAlaLeuValSerThrLeuGly 319
Db      315172 GTATCAATTTATTTGCTCGTGTAAATGAAATGAGTAACTAATAATACCACTTTAGGT 315231
QY      320 ValThrGlyGlnLysAsnLysSerGlyLeuTyrGluAspGlyValArg--LysAspArg 338
Db      315232 CTAGCAGAAAGTCAATAGAAAGTGCATTTTATGCAAGACGAGTAAAGCAAAAAAGAAAA 315291
QY      339 LeuAspAsnAsnAspMetIleAspGlnLeuGluAlaArgGlyLeuArgAlaLysAlaSerGln 358
Db      315292 ATAAAAACGAAAGAAATTAATAGAAAAAATGAAATTAATTAATGAAAAAATAGACAAA 315351
QY      359 LeuAspGluAlaArgArg 364
Db      315352 TTAATTAATTCAAAAAAA 315369

RESULT 4
US-09-974-300-1692
; Sequence 1692, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berke, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10065.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1692
; LENGTH: 1083
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1692

```

```

Alignment Scores:
Pred. No.: 1.07e-89 Length: 1083
Score: 822.00 Matches: 162
Percent Similarity: 68.45% Conservative: 55
Best Local Similarity: 51.10% Mismatches: 100
Query Match: 44.05% Indels: 0
DB: 10 gaps: 0
US-09-921-992-78 (1-372) x US-09-974-300-1692 (1-1083)

```

```

QY      9 ArgArgYSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaProIle 28
Db      24 CSTRCAAAAAAGCGGCTCCCGTTAAAGTGGACCTTTACACATATGGCGCAATACGAAGATC 83
QY      29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48
Db      84 GTCATTCAAAGCATGACCAACACGAACACATGAGCTTGAAGCAACCGTCCCGCAATC 143
QY      49 LysAlaLeuGlnLysValGlyAlaAspIleValAlaArgValSerValProThrMetSpar 68
Db      144 AACAGACTCGCGAGACAGAGATGTCAAATCTCCGCTGCTGCTGATGACAGCGCT 203
QY      69 AlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHis 88
Db      204 GCCGAGCCCATTCACAGATCAAAAGCGGATTCATCCCTCTGCTGCGATTTCT 263
QY      89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
Db      264 TTCACATTAATATTTGCGATTAAGAGCATGCAAGCGGAGCGCGATTAATCCGATTCAT 323
QY      109 ProGlyAsnIleGlyAsnLysGlnLysGlyIleArgMetValAlaAspCysAlaArgAspLys 128
Db      324 CCGGATACATGCGCGCGCGGCAAAAGTTGAAGCGGTGTACAGCGCAAGCAAGAAAG 383
QY      129 AsnIleProIleArgIleGlyValAlaAsnArgIleSerLeuGlnLysAspLeuGlnLys 148
Db      384 GGCATTCGATCCGATCCGATCGGCTCAATGAGGCTCTCTGGAAGAAAGCAATCCTTAGAG 443
QY      149 TyrGlyGluProThrProGlnAlaLeuGluGlnSerAlaMetArgHisValAspHisLeu 168
Db      444 TACGGGTATCCACAGACAGACAGCGCATGTCGCAAAAGCGGCTGACCATTAATTTCT 503
QY      169 AspArgLeuAsnPhaeSparGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
Db      504 GAGATCTCGATTCACAGATATCATCTGTCACATGAAAGGCGTGAATGAACCTGGCG 563
QY      189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
Db      564 ATTGAGGCATATGAAAAAGCGGCTAAAGCTTTCGATTCGCGCTCATTTAGGATCC 623
QY      209 GluAlaGlyGlyAlaArgSerGlyValAlaLysSerAlaIleGlyLeuGlyLeuLeu 228
Db      624 GAATCGGAGAGCGCTTTGCGGATACAGTAAAGCGGCGGCGGCTGCGGCGCATCTCT 683
QY      229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGlnIle 248
Db      684 TCAAAAGGAATCGGCAATATGCGGATTTCTTTAAGCGGAGACCGGCTGGAAGAAAGTA 743
QY      249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPhel 268
Db      744 AAAGTCGCGAAGAGCTCTGTAATCTTTCCGCTGCTGCTCATTCGCGCAACATGAT 803
QY      269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAlaAlaLeuGlu 288
Db      804 TCCTGCCGACTTCGCGCGGATGAAATCATTTGATTCGATTCGCAATGAATTCGAA 863
QY      289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValAlaAsn 308
Db      864 GATTACATCGGCAAAATCAAGCGCGATCAAGATTCGCGCTTCGCGCTGCGCGTAC 923
QY      309 GlyProGlyGluAlaLeuValSerThrLeuGlyAlaThrGlyGlnLys 325

```



OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Alignment Scores:  
Pred. No.: 4,096-82 Length: 9025608  
Score: 808.00 Matches: 164  
Percent Similarity: 64.11% Conservative: 70  
Best Local Similarity: 44.93% Mismatches: 119  
Query Match: 43.30% Indels: 12  
DB: 14 Gaps: 3

US-09-921-992-78 (1-372) x US-10-156-761-1 (1-9025608)

```

QY      8 GlnArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaPro 27
DB      3141566 GAGCCGCGAGAGAGCGGCGGAGTCCAGTCCGGAACCGCGCGGCGGAGGAGCAGCC 3141507
      |||
QY      28 IleAlaValGlnSerMetThrAsnThrArgThrAspValGlnAlaThrValAsnGln 47
DB      3141506 GTCGCGGTCGATGATGACGACGAGCGGTACGTCGACATCGGCGGCGGCGGCGGAG 3141447
      |||
QY      48 IleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAsp 67
DB      3141446 ATCGCGGAGCTCAGCGGCTCCGCGTCCAGATCGTCCGTGCGCTCCGCGGCGGAG 3141387
      |||
QY      68 AlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIle 87
DB      3141386 GACCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3141327
      |||
QY      88 HisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIle 107
DB      3141326 CACTTCCAGCGGAGGAGTACGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3141267
      |||
QY      108 AsnProGlyAsnIleGlyAsn--GluGluArgIleArgMetValValAspCysAlaArg 126
DB      3141266 AACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3141207
      |||
QY      127 AspLysAsnIleProIleArgIleGlyValAlaAsnAlaGlySerLeuGluLysAspLeuGln 146
DB      3141206 GAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3141147
      |||
QY      147 GluLysTyrGlyLysProThrProGlnAlaLeuLeuGlnLysSerLeuAlaAspIle 166
DB      3141146 GAGAACTACGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3141087
      |||
QY      167 HisLeuAspArgLysLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPhe 186
DB      3141086 CTCTTCGAGGAGGAGCAGCTTCGCGGAGCATCAAGATCTCGGTCAAGCAACAGCCGG 3141027
      |||
QY      187 LeuAlaValGlnSerTyrArgLeuAlaLysGlnIleAspGlnProLeuHisLeuGly 206
DB      3141026 GTCATGTCACACCGTCACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3140967
      |||
QY      207 IleThrGluAlaGlyLysAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeu 226
DB      3140966 GTCACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3140907
      |||
QY      227 LeuLeuSerGlnGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGlu 246
DB      3140906 CTCTCTACGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3140847
      |||
QY      247 GluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsn 266
DB      3140846 GAGATCAAGGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3140787
      |||
QY      267 PheIleAlaCysProThrArgSerArgGlnGluPheAspValIleGlyThrValAsnAla 286
DB      3140786 ATCTGTCCTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3140727
      |||
QY      287 LeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyVal 306
DB      3140726 GTCACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3140667
      |||

```

```

QY      307 ValAsnGlyProGlyGlnAlaLeuValSerThrLeuGlyValThrGlyIleAsnLys 326
DB      3140666 GTCACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3140607
      |||
QY      327 SerGlyLeuArgValArgLysAspArgLeuAsnAsnAspMetIleAsp 346
DB      3140606 GGCACAGTCTTCGGAAGGCGGAATCATCAACAGCGGTCCGCGGCGGCGGCGG 3140547
      |||
QY      347 GlnLeuGlnAlaArgIleArgAlaLysAlaSerGlnLeuAspGlnLalaArgArgIleAsp 366
DB      3140546 ACCCTG-----ATCGAGGAGCGGATGAAGATC--- 3140520
      |||
QY      367 ValGlnGlnValGlu 371
DB      3140519 GCCGAGCAGATGAGAG 3140505
      |||

```

# RESULT 7

```

US-09-881-752A-195
Sequence 195, Application US/09881752A
Patent No. US20020115078A1
GENERAL INFORMATION:
APPLICANT: Kleantous, Harold
APPLICANT: Al-Garawi, Amal
APPLICANT: Miller, Charles
APPLICANT: Tomb, Jean-Francois
APPLICANT: Oomen, Raymond P.
TITLE OF INVENTION: Identification of Polynucleotides
TITLE OF INVENTION: Encoding No. US20020115078A1el Helicobacter Polypeptides in t
FILE REFERENCE: 06132/041002
CURRENT APPLICATION NUMBER: US/09/881,752A
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 08/833,457
NUMBER OF SEQ ID NOS: 370
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 195
LENGTH: 1180
TYPE: DNA
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: CDS
LOCATION: (51)...(1127)
US-09-881-752A-195

```

Alignment Scores:  
Pred. No.: 3,026-86 Length: 1180  
Score: 794.00 Matches: 162  
Percent Similarity: 66.38% Conservative: 73  
Best Local Similarity: 45.76% Mismatches: 117  
Query Match: 42.55% Indels: 2  
DB: 10 Gaps: 2

US-09-921-992-78 (1-372) x US-09-881-752A-195 (1-1180)

```

QY      9 ArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaProIle 28
DB      63 AGAGTTAAGACCAACAAATTTTATGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 122
      |||
QY      29 AlaValGlnSerMetThrAsnThrArgThrAspValGlnAlaThrValAsnGlnIle 48
DB      123 AGCAGCGCAACATGACTTTTGAACAAACCGCTGATTTGAAACCACTTAAACCAATTT 182
      |||
QY      49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
DB      183 GACAGACATCAACATCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 242
      |||
QY      69 AlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHis 88
DB      243 GCTCTAGCCTTAAAGATTTGAAAAAGTGTCCCTTGCCTTTAAATCGGTGATATTCAT 302
      |||
QY      89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
      |||

```

```

Db      303 TTCCATTATAATTCGGCTC--ATTGCCGCTCAAGCGTGCATCGCATTAAC 359
      109 ProGlyAsnIleGlyAsnGluGluArgIleArgMetValAlaSPCySaIaArgSPlys 128
      360 CCCGAAACATCGGCTGTAAGAGACATCAAGCGGTGTGATGCTTGAAGAAAA 419
      129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluYAspPleuGlnGlu 148
      420 AACATTCCTATTAAGATTCGGCATATCGGTGAGTTAGAAACGATTTGATGATTA 479
      149 TyrGlyGluProThrProGlnAlaLeuGlnSerAlaMetArgHisValAspHisLeu 168
      480 TACGGA---CCACCCGAAAGAGCATGTGATAAGCCGTTGTATACGCCAACTTTA 536
      169 AsparGleuAsnPheAspGlnPheIysValSerValIysAlaSerAspValPheLeu 188
      537 GAAATTTGATTTAGCAATTTTAAATTTCTTTAAAGAGGAGATGATTCGCACC 596
      189 ValGluSerTyrArgLeuLeuAlaIysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
      597 ATAGAGCTTACAGATCTTCGCCCTTCCTGATCTATCCCTTCATTTGGGGTACG 656
      209 GluAlaGlyGlyAlaArgSerGlyValAlaIysSerAlaIleGlyLeuGlyLeuLeu 228
      657 GAGCGGGGAATCTTTAGCTCCAGTACAAATCCGCTATGCGCTTAGGGGGCTTTTA 716
      229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaIleAspProValGluGlu 248
      717 ATGGAGGGCATTTGGGATTCGATTCGATTCACACAGGGGAATAGAAAATATATC 776
      249 LysValGlyPheAspIleLeuIysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
      777 AAAGTGGCCAGACAAATTTACGCATACCGCGCGTTGAAGAGGATTTATTTGAT 836
      269 AlaCysProThrCysSerArgIleGluPheAspValIleGlyThrValAsnAlaLeuGlu 288
      837 TCTTGCCCATCTTCGGCGCATTTGAGCCAAATTTAGTGCATATGCGCATCAAGTAGA 896
      289 GlnArgLeuGlnAspIleIleThrProMetAspValSerIleIleGlyCysValAlaAsn 308
      897 AAACGCTTAAGCCACATTAACCCCTTACATTTAGCATTTAGGTGCGTGTGAT 956
      309 GlyProGlyIleAlaIleValSerThrLeuGlyValThrGlyCysAlaIysSerGly 328
      957 GCTTGGGGAAGCAAGCATGACATGCGCATCGCTTTGGGAATGCGACGGTTTG 1016
      329 LeuTyrGluAspGlyValArgIysAspArgLeuAspAsnAspMetIleAspGlnLeu 348
      1017 ATCATTAAGAGCTTAAGTCAATTCACAACTGCTGAAGAGATTTATTTGAACTTT 1076
      349 GluAlaArgIleArgAlaIysAlaSerGlnLeuAspGluAla 362
      1077 GTGATAGAAAGTGAATTTAGCTAAGAAAGAGAAAAAAGT 1118

```

```

      PRIOR FILING DATE: 2001-08-02
      NUMBER OF SEQ ID NOS: 15109
      SEQ ID NO 1536
      LENGTH: 1155
      TYPE: DNA
      ORGANISM: Streptomyces avermitilis
      FEATURE:
      NAME/KEY: CDS
      LOCATION: (1)..(1155)
      US-10-156-761-1636

Alignment Scores:
Pred. No.: 2,37e-85 Length: 1155
Score: 786.50 Matches: 164
Percent Similarity: 63.69% Conservative: 71
Best Local Similarity: 44.44% Mismatches: 121
Query Match: 42.15% Indels: 13
Gaps: 4

US-09-921-992-78 (1-372) x US-10-156-761-1636 (1-1155)
      6 Profile---GlnArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAsp 24
      43 CCGATCGCCGCAACCCGTGCTCCCGGCAGATCCAGTCCGACCGCGTGGCGGCGGC 102
      25 GlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAspValGluAlaThr 44
      103 ACGGCCCGCGGTGGTGCATGATGACGACGCGGTACGTCGTCGATCGGCGGCGCACG 162
      45 ValAsnGlnIleIysAlaLeuGluArgValGlyAlaAspIleValArgValSerValPro 64
      163 CTCAGCAGATCGCGGAGACTACCGCGTCCGCGTCCAGATCGTCCGCGTCCGCGCC 222
      65 ThrMetAspAlaAlaGluAlaPheIysLeuIleIysGlnGlnValAsnValProLeuVal 84
      223 ACGCAGCAGCAGCGGAGCGCCCTCGGCTATCGCCGCAAGTCGACAGATCCCGTGTGC 282
      85 AlaAspIleHisPheAspTyrArgIleAlaLeuIysValAlaGluTyrGlyValAspCys 104
      283 GCGGACATCCACTTCACCGAAGTACGTTCGCGCGGCGATCGACGCGCGTCCGCGCG 342
      105 LeuArgIleAsnProGlyAsnIleGlyAsn--GlnGluArgIleArgMetValAlaAsp 123
      343 GTCCGGGTGMAACCGCGGACATCAAGCATTCGACGACGACGACGACGATCGCGCGC 402
      124 CysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGlnLys 143
      403 GCGGCCAAGAGACAGCGCAGCGCATTCGACATCGCGCTCAACGCGGCGTCTGACCGG 462
      144 AspLeuGlnGlnLysTyrGlyGluProThrProGlnAlaLeuGlnSerAlaMetArg 163
      463 CGGTGCTGGAAGATACGGACAGGCCACCCCGAGGCGCTCTGATCGTGGCGGTG 522
      164 HisValAspHisLeuAspArgLeuAsnPheAspGlnPheIysValSerValIysAlaSer 183
      523 GAGGGGTGCTCTTCGAGAGACAGCACTTCGCGCATCAAGATCTCGTCAAGACAC 582
      184 AspValPheLeuAlaValGluSerTyrArgLeuLeuAlaIysGlnIleAspGlnProLeu 203
      583 GACCGGTGCTCATGCTCAACGCTACCGCAGCTGGGCGGCGACATGCGACTACCGCGTG 642
      204 HisLeuGlyIleThrGluAlaGlyIleAlaArgSerGlyAlaValIysSerAlaIleGly 223
      643 CACCTCGCGGTACCGAGCGCGCGCGCTTCGACGACGATCAATCGCGCGCTGCC 702
      224 LeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaIleAsp 243
      703 TTCGGCGCTCTCTGACGACGAGGCAATCGCGCACAGATCCGCTGCTGAGCGCGCC 762
      244 ProValGluGlnIleIysValGlyPheAspIleLeuIysSerLeuArgIleArgSerArg 263
      763 CCGGTGAGAGATCAAGGTCCGATCCAGATCTGGAATCGCTGGCGTCAAGGAGCGG 822

```

```

Db      73  TCGGTCCAGTCGATGACCAACCACCAACCCAGCACTCAAGGCAACCTTCGAACAGATC 132
QY      49  LysAlaLeuGluArgValIGlyAlaAspIleValArgValSerValProthrMetAspAla 68
Db      133 GCACAGTTGACAGCCACCGCGTTGGACATGTCGCGCTTCGCCCAAGACGTGTGAT 192
QY      69  AlAGluAlaPheLysIleuIleLysGlnGluValAsnValProLeuValAlaAspIleHis 88
Db      193 GCGGAAGCACTGCGCATCTCGCAAGAAAGATCTCCGATCCCAAGATGCGAGATATCCAC 252
QY      89  PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAlaSpCysLeuArgIleAsn 108
Db      253 TTCACGCCCAAGACATCTTCGCGGCACATGACAGTTGCGCCCGCGCTTCGHTGCAC 312
QY      109 ProGlyAsnIleGlyAsn---GluGluArgIleArgMetValAlaAspCysAlaIleArgAsp 127
Db      313 CCAGGCAACATCAAGAAATTCGATGTGTCGGCGGTAAAGAAATACCAAAAGCTGCAAGCGAT 372
QY      128 LysAsnIleProIleArgIleGlyValAlaAsnAlaGlySerLeuGluLysAspLeuGlnGlu 147
Db      373 GCCGGAATTCCAATTCGATATGTGTGCACGCGCGGATTCCTGTGAAGCGCAATCTGCAC 432
QY      148 LysTyr---GlyGluProThrProGlnAlaLeuLeuGluSerIleMetArgHisValAsp 166
Db      433 AATATACCAGGCAAAACCCACCAAGAGCTCTGTGGAATCCCAATGTGGAGCCGCC 492
QY      167 HisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPhe 186
Db      493 CTGTTGTAAGAGCACGCTTCGCGCGACATCCGCAATCTCTGTGAAGCATCCGACCAAGTA 552
QY      187 LeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProIleuHisLeuGly 206
Db      553 CTCATGTCGTGAAGCCCTACCGCCAGCTCGCTGACAAAGGACTACCCACTGCACCTCGGT 612
QY      207 IleThrGluAlaGlyLysAlaArgSerGlyValaValLysSerAlaIleGlyLeuGlyLeu 226
Db      613 GTTACGAAGCTGCTGCCAAGTTCAGGAACATCAAGTCTTCGTGACATTCGGCGGT 672
QY      227 LeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaIleAspProValGlu 246
Db      673 CTGCTGTCCACAGGCAATCGGCACACTATCCGTGTCTCTTCTGTGCTACCCAGTGGA 732
QY      247 GluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgIlyIleAsn 266
Db      733 GAAATTCAGGTGGCGGACCAAGATTCGCAGTCCCTCAACCTGCGCCCAAGCAAGTGGAA 792
QY      267 PheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAla 286
Db      793 ATCGTGTCTGCGCCATCATGTGGCGCGCGACAGTCGATGCTACTTCACTTGTGAAGAA 852
QY      287 LeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysVal 306
Db      853 GTACACCGAAGCACTGACGCGGCAATGAAAGTTCCACTGCGGCTCGCTGATGGGTTCGTT 912
QY      307 ValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyLysAsnLys 326
Db      913 GTTAAACGCGCCAGTGAAGCTCGCACGCTGCACTCGGTGTGATCCGGTAAACGCGAAG 972
QY      327 SerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAspAsnAspMetIleAsp 346
Db      973 GGCACGATCTTGTCAAGGCGCAATCATCAAGACTGTCCAGATCCCAATCTCGGA 1032
QY      347 GlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGlnAlaArgIle 365
Db      1033 ACCCTC-----ATCGAAGAAAGCAATCGTATC 1059

RESULT 10
US-09-738-626-1/C
: Sequence 1, Application US/09738626
: Publication No. US20020197605A1
: GENERAL INFORMATION:
: APPLICANT: NAKAGAWA, SATOSHI
: APPLICANT: MIZOGUCHI, HIROSHI

```



## Alignment Scores:

Pred. No.: 4,2e-81 Length: 1164  
 Score: 751.50 Matches: 156  
 Percent Similarity: 60.27% Conservative: 64  
 Best Local Similarity: 42.74% Mismatches: 134  
 Query Match: 40.27% Indels: 11  
 DB: 10 Gaps: 2

US-09-921-992-78 (1-372) x US-09-712-363-110 (1-1164)

```

OY 9 ArgArglySerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaProIle 28
Db CGCGCGCCGACCGCTGTCAGCTGTCGCAACGTCGCGGACGTGCGACATCCGCTC 111
OY 29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48
Db TCGGTGCAATCGATGTCACACCAACCAACCACTCAACTCGACATTCGACACATATC 171
OY 49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
Db 172 GCCGAGTGCACCGCGCGGATGCGACATGTCGCGGTGCGCTGCGCGCCAGAGGAC 231
OY 69 AlaGluAlaPheLysLeuIleTyrGlnGlnValAsnValProLeuValAlaAspIleHis 88
Db 232 CGCGAGCGCTGGCGCGGATCGCGCGACAGCAGATCCCGGTAGTCGCGGACATACAT 291
OY 89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAlaAspCysLeuArgIleAsn 108
Db 292 TTCACGCGCGACATATTCGCGCGCATGCGCGGATGTCGCGCGGCGCGGTCAC 351
OY 109 ProGlyAsnIleGlyAsn---GluGluArgIleArgMetValValAspCysAlaAsp 127
Db 352 CCGGCGACATCAAGAGATTTGACGGCGGCGGTGCGAGTCCCAAGCGCGCGTGC 411
OY 128 LysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnIle 147
Db 412 GCCGCGATCCGATCGCAATCGGTGTCACAGCGCGTTCGTGACAAAGCGTTCATGAG 471
OY 148 LysTyrGlyGluProThrProGlnAlaLeuGluSerAlaMetArgHisValAspHis 167
Db 472 AAGTATGCGCAACGCCCGCGCGGCGGTGACGCTGCGCGCGCTGCGGAGCGTTC 531
OY 168 LeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspAlaPheLeu 187
Db 532 TTGAGAGACATGCGCTTCGTCATCAAGATCAGCGTCAAGCAACCGCGGTGCG 591
OY 188 AlaValGlnSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIle 207
Db 592 ATGTCGCGCGCTACGAGCTGCTGTCACGCGTGCAGCTTACCACTGCACTCGGTGC 651
OY 208 ThrGluAlaGlyLysAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeu 227
Db 652 ACCGAGCGCGCGCTTCCTTCCAGGCGCACATCAAGTCCGCGTTCGCGCGCTG 711
OY 228 LeuSerGlnGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGlu 247
Db 712 CTGTGCGCGGATGACGACACATCCGCGGTGTCGCGCGCGCGCGGTGCAAGAA 771
OY 248 IleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgIlyIleAsnHe 267
Db 772 GTCAAGTGGCAATCAGCTTCTGAGTCTGTAACCTGGCGCGCGCTTCGTCAGATC 831
OY 268 IleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeu 287
Db 832 GTGTCTTCCCGCTGCGCGGTGCGCGCAAGTGCAGCTTACACCTCGCCCAACGAGTA 891
OY 288 GluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValAla 307
Db 892 ACCGCGCGCGCTGATGCTCATGTCGCGCGGTGCGCGGTGCGGTGCTGCTGCTC 951
OY 308 AsnGlyProGlyLysAlaLeuValSerThrLeuGlyValThrGlyLysLysSer 327
  
```

```

Db 952 ATGTCGCGGCTAAGACAGCTGAGCGGACCTGCGGCGGTCCGCAACGCAAGGT 1011
OY 328 GlyLeuTyrGlyLysAspGlyValArgLysAspArgLeuAsnAsnAspIleAspGln 347
Db 1012 CAGATCTTGTACGGCGGCGAATGATCAAGACCGTCCCGAAGCACAGATCTGCGAGC 1071
OY 348 LeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAlaArgArgIleAspVal 367
Db 1072 CTC-----ATCGAGAGGCGATGCGGCTGCGCGCC 1101
OY 368 GlnGlnValGluLys 372
Db 1102 GAATGGCGGAGCA 1116
  
```

## RESULT 12

```

US-09-921-992-1
; Sequence 1, Application US/09921992
; Patent No. US20020069426A1
; GENERAL INFORMATION:
; APPLICANT: Boronati, Albert;
; APPLICANT: Campos, Narciso;
; APPLICANT: Rodriguez-Concepcion, Manuel;
; APPLICANT: Rohmer, Michel;
; APPLICANT: Seeman, Myriam;
; APPLICANT: Valentin, Henry E.;
; APPLICANT: Venkatesh, Tyamagondlu V.;
; APPLICANT: Venkatesh, Mylavaram;
; TITLE OF INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes
; FILE REFERENCE: 16516.107/35-21(51897)US
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/223,483
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 85
; SEQ ID NO 1
; LENGTH: 2520
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (154)..(2376)
US-09-921-992-1
  
```

## Alignment Scores:

Pred. No.: 1.66e-45 Length: 2520  
 Score: 462.00 Matches: 105  
 Percent Similarity: 55.88% Conservative: 47  
 Best Local Similarity: 38.60% Mismatches: 86  
 Query Match: 24.76% Indels: 34  
 DB: 9 Gaps: 5

US-09-921-992-78 (1-372) x US-09-921-992-1 (1-2520)

```

OY 9 ArgArglySerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaProIle 28
Db 403 AGAAGGAGACCTCTACGATGTTATGTTGAATCTCGCGCTTGAAGCAACATCGATA 462
OY 29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48
Db 463 AGGATTCACAGATGACTTCTGATTCACAAAGATATTTACTGGAACCTGTTGATGAGGT 522
OY 49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
Db 523 ATGAGATAGCGGATTAAGAGATGATATTTGTAAGATTAATCTTCAAGCAAGCAAGAG 582
OY 69 AlaGluAlaPhe-----LysLeuIleLysGlnGlnValAsnValProLeu 83
Db 583 GCGGATCGCGCTTGAATTAAGATTAAGATTAATCTTCAATTAATACCGCTG 642
OY 84 ValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAsp 103
Db 643 GTTCACATATTCATTTTTCGCGCGCTTACGAGTTCGTCGATGATCC---TTTGAC 699
  
```







Query Match: 18.17% Indels: 24  
 DB: 9 Gaps: 5  
 US-09-921-992-78 (1-372) x US-09-921-992-22 (1-670)

QY 133 ArgIleGlyValAlaSerLeuGlu-LysAspLeuGlnGluLysTyrGlyGluPr 152  
 |||  
 15 AGGATCGGCGTTAACCATGCTCTCTCCGAAAGAAATGC-----TTTACCTACTTTT 68  
 QY 152 oThrProGlnAlaLeuLeuGlnSerAlaMetArgHisValAspHisLeuAspArgLeuAs 172  
 |||  
 69 TACCCCGAGGGCATGTCATGCGATCGCCCTGGAATTCATAAATTGTGTGCTCTAGA 128  
 QY 172 nPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaValGluSerTy 192  
 |||  
 129 TTTCCGCAACCTAGTCGTTCATGAAAGCCTCCCGGTACCGGTAAATGTTGGCGGCTA 188  
 QY 192 rArgLeuLeuAlaLysGlnIleAspGln-----ProLeuHisLeuGlyI 207  
 |||  
 189 TCGCCTCATGCTGTAAGCATGATGAGCATGTGGCATGATTAATCCCTCATCTAGGGGT 248  
 QY 207 eThrGlnAlaGlyLysAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeu 227  
 |||  
 249 TACCGAAGCCGCGGATGGGGAATATGCGCGCATTAATCCACCGCTGCAATGCGACCT 308  
 QY 227 uLeuSerGlnGlyIleGlyAspThrLeuArgValSerLeuAlaIleAspProValGluG 247  
 |||  
 309 TTTAGCTGATGCGATTGCGCATTAATCCCGGTATCCCTCACGAAACCCCGAAAGA 368  
 QY 247 uIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPh 267  
 |||  
 369 AATTCCCGTTGCTACAGCATTTCTCCAGCGCTGGTTCGGAAGAACCATGTTGGAATA 428  
 QY 267 eIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLe 287  
 |||  
 429 TGTGGCTGTCTCTCTCTGTCGCGCATGTTTCATAC-----TT 467  
 QY 287 uGlnGlnArgLeuGlnAspIle-----IleThrProMetAsp-ValS 301  
 |||  
 468 GGAAGACGTGTTCATGATGATCCGAGATGCCACTAAACATCTAACGGTTTAACTTTGCG 527  
 QY 301 eIleIleGlyCysValAlaSerGlyProGlyGlnAlaLeuValSerThrLeuGlyValT 321  
 |||  
 528 CCGTCATGCGCTGATGTCATGATGCGCCCGCGGCATGCGCATGCGCATGATGCTATG 587  
 QY 321 hTrGlyGlnLysLysSerGly 328  
 |||  
 588 TGGGT-----AAACAAGCGGT 604  
 DB 588 TGGGT-----AAACAAGCGGT 604

RESULT 15  
 US-09-921-992-23  
 ; Sequence 23. Application US/09921992  
 ; Patent No. US20020069426A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boronati, Albert;  
 ; APPLICANT: Campos, Narciso;  
 ; APPLICANT: Rodriguez-Concepcion, Manuel;  
 ; APPLICANT: Rohmer, Michel;  
 ; APPLICANT: Seeman, Myriam;  
 ; APPLICANT: Valentin, Henry E.;  
 ; APPLICANT: Venkatesh, Tyamagondlu V.;  
 ; TITLE OF INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes  
 ; FILE REFERENCE: 16516.107/25-21(51897)US  
 ; CURRENT APPLICATION NUMBER: US/09/921,992  
 ; PRIOR FILING DATE: 2001-08-06  
 ; NUMBER OF SEQ ID NOS: 85  
 ; SEQ ID NO 23  
 ; LENGTH: 596  
 ; TYPE: DNA  
 ; ORGANISM: Zea mays

FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: (1..596)  
 ; OTHER INFORMATION: unsure at all n locations  
 US-09-921-992-23

Alignment Scores:  
 Pred. No.: 5.82e-30 Length: 596  
 Score: 326.00 Matches: 78  
 Percent Similarity: 58.62% Conservative: 41  
 Best Local Similarity: 38.42% Mismatches: 62  
 Query Match: 17.47% Indels: 22  
 DB: 9 Gaps: 4

US-09-921-992-78 (1-372) x US-09-921-992-23 (1-596)

QY 133 ArgIleGlyValAlaSerLeuGlu-LysAspLeuGlnGluLysTyrGlyGluPr 152  
 |||  
 15 AGGATCGGCGTTAACCATGCTCTCTCCGAAAGAAATGC-----TTTACCTACTTTT 68  
 QY 152 oThrProGlnAlaLeuLeuGlnSerAlaMetArgHisValAspHisLeuAspArgLeuAs 172  
 |||  
 69 TACCCCGAGGGCATGTCATGCGATCGCCCTGGAATTCATAAATTGTGTGCTCTAGA 128  
 QY 172 nPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaValGluSerTy 192  
 |||  
 129 TTTCCGCAACCTAGTCGTTCATGAAAGCCTCCCGGTACCGGTAAATGTTGGCGGCTA 188  
 QY 192 rArgLeuLeuAlaLysGlnIleAspGln-----ProLeuHisLeuGlyI 207  
 |||  
 189 TCGCCTCATGCTGTAAGCATGATGAGCATGTGGCATGATTAATCCCTCATCTAGGGGT 248  
 QY 207 eThrGlnAlaGlyLysAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeu 227  
 |||  
 249 TACCGAAGCCGCGGATGGGGAATATGCGCGCATTAATCCACCGCTGCAATGCGACCT 308  
 QY 227 uLeuSerGlnGlyIleGlyAspThrLeuArgValSerLeuAlaIleAspProValGluG 247  
 |||  
 309 TTTAGCTGATGCGATTGCGCATTAATCCCGGTATCCCTCACGAAACCCCGAAAGA 368  
 QY 247 uIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPh 267  
 |||  
 369 AATTCCCGTTGCTACAGCATTTCTCCAGCGCTGGTTCGGAAGAACCATGTTGGAATA 428  
 QY 267 eIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLe 287  
 |||  
 429 TGTGGCTGTCTCTCTCTGTCGCGCATGTTTCATAC-----TT 467  
 QY 287 uGlnGlnArgLeuGlnAspIle-----IleThrProMetAsp-ValS 301  
 |||  
 468 GGAAGACGTGTTCATGATGATCCGAGATGCCACTAAACATCTAACGGTTTAACTTTGCG 527  
 QY 301 eIleIleGlyCysValAlaSerGlyProGlyGlnAlaLeuValSerThrLeuGlyValT 321  
 |||  
 528 CCGTCATGCGCTGATGTCATGATGCGCCCGCGGCATGCGCATGCGCATGATGCTATG 587  
 QY 321 hTrGly 322  
 |||  
 588 TGGGT 592  
 DB 588 TGGGT 592

Search completed: August 15, 2003, 10:15:58  
 Job time : 27217 secs

\_\_\_\_\_

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2003, 09:54:00 ; Search time 41 seconds  
(without alignments)  
872.555 Million cell updates/sec

Title: US-09-921-992-78

Perfect score: 1866

Sequence: 1 MHNOAPIORRKRSTRIVGNV.....RAKASQDEARRIDYQVEK 372

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.76:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1866	100.0	372	1 S23058	gcpe protein - Esc
2	1866	100.0	372	2 A91051	hypothetical prote
3	1866	100.0	372	2 E85895	hypothetical prote
4	1831	98.1	372	2 AB0822	gcpe protein (prot
5	1622.5	87.0	375	2 AG0350	probable acetyltra
6	1621.5	86.9	376	2 F82283	gcpe protein VC075
7	1603.5	85.9	376	1 H64063	gcpe protein - Hae
8	1372.5	73.6	371	2 F83171	conserved hypotnet
9	1282	68.7	368	2 E84963	gcpe protein (limpo
10	1042.5	55.9	349	2 H87354	gcpe protein (limpo
11	887	47.5	343	2 G97121	gcpe protein (limpo
12	882.5	47.3	367	2 AB3825	peptidoglycan acet
13	868	46.5	368	2 A11254	probable peptidogly
14	859.5	46.1	377	2 C69935	peptidoglycan acet
15	828	44.4	357	2 F70433	gcpe protein - Aqu
16	815.5	43.7	384	2 T35407	conserved hypotnet
17	794	42.6	359	1 A64598	hypothetical prote
18	751.5	40.3	387	2 F70886	probable gcpe prot
19	712.5	38.2	392	2 G87106	conserved hypotnet
20	706	37.8	344	2 D72321	gcpe protein - The
21	690	37.0	357	2 AB1339	gcpe protein homol
22	618.5	33.1	403	2 S77159	gcpe protein - Syn
23	613.5	32.9	408	2 AF2118	hypothetical prote
24	588	31.5	429	2 D75526	probable gcpe prot
25	543.5	29.1	421	2 D81098	gcpe protein - Del
26	527	28.2	421	2 H81843	conserved hypotnet
27	527	28.2	421	2 H81843	gcpe protein (limpo
28	527	28.2	421	2 H81843	gcpe protein (limpo
29	509	27.3	437	2 AH3285	gcpe protein (limpo

#### ALIGNMENTS

##### RESULT 1

S23058  
gcpe protein - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 01-Mar-2002

C:Accession: S23058; B65028

R:Baker, J.; Franklin, D.B.; Parker, J.

FEMS Microbiol. Lett. 94, 175-180, 1992

A:Title: Sequence and characterization of the gcpe gene of Escherichia coli.

A:Reference number: S23057

A:Accession: S23058

A:Molecule type: DNA

A:Residues: 1-372 <BAK>

A:Cross-references: EMBL:X64451; NID:g41540; PID:g41542

A:Experimental source: strain K-12

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:9742617; PMID:9278503

A:Accession: B65028

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Residues: 1-372 <BLAT>

A:Cross-references: GB:A000338; GB:U00096; NID:g178862; PID:AAC75568.1; PID:g17888

A:Experimental source: strain K-12, substrain MGI655

C:Gene(s):

A:Gene: gcpe

C:Superfamily: gcpe protein

Query Match 100.0% Score 1866; DB 1; Length 372;

Best Local Similarity 100.0% Pred. No. 1.3e-121;

Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy	1	MHNOAPIORRKRSTRIVGNVPIGDAPIAVOSMTNRTDVEATYNOIKALERGADIVR	60
Db	1	MHNOAPIORRKRSTRIVGNVPIGDAPIAVOSMTNRTDVEATYNOIKALERGADIVR	60
Oy	61	VSVPYMAAEAFKLIKQOVNPLVADHEDYRIALKAVEYGVDCRLINPNGNEERTIM	120
Db	61	VSVPYMAAEAFKLIKQOVNPLVADHEDYRIALKAVEYGVDCRLINPNGNEERTIM	120
Oy	121	VVDCARDKNIPIRIGVNAAGSLKDLQKRYGPTQALLESAMRHVDHLDRLNFOQFVSV	180
Db	121	VVDCARDKNIPIRIGVNAAGSLKDLQKRYGPTQALLESAMRHVDHLDRLNFOQFVSV	180
Oy	181	KASDVLAESRYLLAKQIDOPHLGITTEAGGASGAVKSAIGLGLLSGIGDTLAVSL	240
Db	181	KASDVLAESRYLLAKQIDOPHLGITTEAGGASGAVKSAIGLGLLSGIGDTLAVSL	240
Oy	241	AADPEETIKYGFOLKLRIRSRKGINFIACPTGSRQGFVDYIGVNALEQRLEDITPMVY	300
Db	241	AADPEETIKYGFOLKLRIRSRKGINFIACPTGSRQGFVDYIGVNALEQRLEDITPMVY	300

Db 241 AADVEEIKVGFILASRIRSRGNIACPTCSROEFVIGTVNALEORLEDTITPMV 300  
 QY 301 SIICGVNNGEALVSTLGVGTGNGKSGLYEDGVKRDLDNNMDIDLEARIKASQLD 360  
 Db 301 SIICGVNNGEALVSTLGVGTGNGKSGLYEDGVKRDLDNNMDIDLEARIKASQLD 360  
 QY 361 EARRIDVOOVER 372  
 Db 361 EARRIDVOOVER 372

## RESULT 2

A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gene  
 A:Reference number: A91051  
 A:Accession: A91051  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-372 <HAV>  
 A:Cross-references: GB:BA000007; PIDN:BA836800.1; PID:G13362847; GSPDB:GN00154  
 A:Experimental source: strain 0157:H7, substrain RMD 0509552  
 A:Gene: ECs3377  
 C:Superfamily: gcpe protein

Query Match 100.0%; Score 1866; DB 2; Length 372;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-121;  
 Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHNQAPIORRSTRIYGVNPIGDGAPIAVOSMTNRTTVEATVNOIKALERGADIVR 60  
 Db 1 MHNQAPIORRSTRIYGVNPIGDGAPIAVOSMTNRTTVEATVNOIKALERGADIVR 60  
 QY 61 VSVPTMDAAEAFKLIKQOVNPLVADHFDYRIALKAAYEGVDCLRINPNIENRIRM 120  
 Db 61 VSVPTMDAAEAFKLIKQOVNPLVADHFDYRIALKAAYEGVDCLRINPNIENRIRM 120  
 QY 121 VVDCARDKNIPIRIGVNGSLKEDLOEKYGEPTPOALLSARHVDHLRLNFDQFVSV 180  
 Db 121 VVDCARDKNIPIRIGVNGSLKEDLOEKYGEPTPOALLSARHVDHLRLNFDQFVSV 180  
 QY 181 KASDVFLAVESYRLAKQIDOPPLHGLITTEAGARSGAVKSAIGLLSGIGDTLRVSL 240  
 Db 181 KASDVFLAVESYRLAKQIDOPPLHGLITTEAGARSGAVKSAIGLLSGIGDTLRVSL 240  
 QY 241 AADVEEIKVGFILASRIRSRGNIACPTCSROEFVIGTVNALEORLEDTITPMV 300  
 Db 241 AADVEEIKVGFILASRIRSRGNIACPTCSROEFVIGTVNALEORLEDTITPMV 300  
 QY 301 SIICGVNNGEALVSTLGVGTGNGKSGLYEDGVKRDLDNNMDIDLEARIKASQLD 360  
 Db 301 SIICGVNNGEALVSTLGVGTGNGKSGLYEDGVKRDLDNNMDIDLEARIKASQLD 360  
 QY 361 EARRIDVOOVER 372  
 Db 361 EARRIDVOOVER 372

## RESULT 3

E85895  
 Hypothetical protein gcpe [imported] - *Escherichia coli* (strain 0157:H7, substrain EDL93)  
 C:Species: *Escherichia coli*  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C:Accession: E85895  
 R:Perera, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 Miller, L.; Grothbeck, E.J.; Davids, N.W.; Lim, A.; Dimantanta, E.; Potamoustas, K.; Apodaca,

Nature 409, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
 A:Reference number: A85480; MUID:21074935; PMID:11206551  
 A:Accession: E85895  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-372 <STO>  
 A:Cross-references: GB:AE005174; NID:G12516909; PIDN:AA657625.1; GSPDB:GN00145; UMGSP:  
 A:Experimental source: strain 0157:H7, substrain EDL93  
 A:Gene: gcpe  
 C:Superfamily: gcpe protein

Query Match 100.0%; Score 1866; DB 2; Length 372;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-121;  
 Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHNQAPIORRSTRIYGVNPIGDGAPIAVOSMTNRTTVEATVNOIKALERGADIVR 60  
 Db 1 MHNQAPIORRSTRIYGVNPIGDGAPIAVOSMTNRTTVEATVNOIKALERGADIVR 60  
 QY 61 VSVPTMDAAEAFKLIKQOVNPLVADHFDYRIALKAAYEGVDCLRINPNIENRIRM 120  
 Db 61 VSVPTMDAAEAFKLIKQOVNPLVADHFDYRIALKAAYEGVDCLRINPNIENRIRM 120  
 QY 121 VVDCARDKNIPIRIGVNGSLKEDLOEKYGEPTPOALLSARHVDHLRLNFDQFVSV 180  
 Db 121 VVDCARDKNIPIRIGVNGSLKEDLOEKYGEPTPOALLSARHVDHLRLNFDQFVSV 180  
 QY 181 KASDVFLAVESYRLAKQIDOPPLHGLITTEAGARSGAVKSAIGLLSGIGDTLRVSL 240  
 Db 181 KASDVFLAVESYRLAKQIDOPPLHGLITTEAGARSGAVKSAIGLLSGIGDTLRVSL 240  
 QY 241 AADVEEIKVGFILASRIRSRGNIACPTCSROEFVIGTVNALEORLEDTITPMV 300  
 Db 241 AADVEEIKVGFILASRIRSRGNIACPTCSROEFVIGTVNALEORLEDTITPMV 300  
 QY 301 SIICGVNNGEALVSTLGVGTGNGKSGLYEDGVKRDLDNNMDIDLEARIKASQLD 360  
 Db 301 SIICGVNNGEALVSTLGVGTGNGKSGLYEDGVKRDLDNNMDIDLEARIKASQLD 360  
 QY 361 EARRIDVOOVER 372  
 Db 361 EARRIDVOOVER 372

## RESULT 4

AB0822  
 Gcpe protein (protein E) [imported] - *Salmonella enterica* subsp. *enterica* serovar Typh  
 C:Species: *Salmonella enterica* subsp. *enterica* serovar Typh  
 A:Note: this species has also been called *Salmonella typhi*  
 C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
 C:Accession: AB0822  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Church  
 th, T.; Connerton, P.; Cronin, A.; Davids, P.; Davies, R.M.; Dowd, L.; White, N.; Farr  
 S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,  
 A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* se  
 A:Reference number: AB0502; MUID:21534947; PMID:11677608  
 A:Accession: AB0822  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-372 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD02726.1; PID:G16503738; GSPDB:GN00176  
 A:Gene: STY2768  
 C:Superfamily: gcpe protein

Query Match 98.1%; Score 1831; DB 2; Length 372;  
 Best Local Similarity 98.1%; Pred. No. 3.5e-119;  
 Matches 365; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```

QY 1 MHNOAPIORRKRSTRIYGVNPIGDCAPIAVOSMTNTRTDDVATVNOIKALERGVADIVR 60
DB 1 MHNOAPIORRKRSTRIYGVNPIGDCAPIAVOSMTNTRTDDVATVNOIKALERGVADIVR 60
QY 61 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAVEYGVDCLRINPNGINEERIRIM 120
DB 61 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAVEYGVDCLRINPNGINEERIRIM 120
QY 121 VVDCARDKNIPIRIGVNGSLEKDOEKYGEPTPOALLESAMRHVDHLDRNFDOFKVSV 180
DB 121 VVDCARDKNIPIRIGVNGSLEKDOEKYGEPTPOALLESAMRHVDHLDRNFDOFKVSV 180
QY 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARSAGVSAIGLILSEGIGDTLRVSL 240
DB 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARSAGVSAIGLILSEGIGDTLRVSL 240
QY 241 AADPVEEIKVGFIDILSLRIRANGINFACPTCSROEFVIGTVNMLEQRLEDITIPMDV 300
DB 241 AADPVEEIKVGFIDILSLRIRANGINFACPTCSROEFVIGTVNMLEQRLEDITIPMDV 300
QY 301 SIICGVNNGPGEALVSTLGVTGNGKSGLYEDGVRRDRLNMDIDLEARIKASQOLD 360
DB 301 SIICGVNNGPGEALVSTLGVTGNGKSGLYEDGVRRDRLNMDIDLEARIKASQOLD 360
QY 361 EARRIDVQOYER 372
DB 361 EARRIDVQOYER 372

```

## RESULT 5

```

AG0350
probable acetyltransferase aarc (imported) - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: AG0350
R:Parthill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.
demo-Tarrega, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
M. J. Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; M01D:21470413; PMID:11586360
A:Accession: AG0350
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-375 <KUR>
A:Cross-references: GB:AL590842; PIDN:q15980846; GSPDB:GN00175
A:Gene: aarc
C:Genetics:
C:Superfamily: gcpe protein

```

```

Query Match 87.0%; Score 1622.5; DB 2; Length 375;
Best Local Similarity 86.8%; Pred. No. 9.1e-105;
Matches 323; Conservative 19; Mismatches 29; Indels 1; Gaps 1;
QY 1 MHNOAPIORRKRSTRIYGVNPIGDCAPIAVOSMTNTRTDDVATVNOIKALERGVADIVR 60
DB 1 MHNOAPIORRKRSTRIYGVNPIGDCAPIAVOSMTNTRTDDVATVNOIKALERGVADIVR 60
QY 61 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAVEYGVDCLRINPNGINEERIRIM 120
DB 61 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAVEYGVDCLRINPNGINEERIRIM 120
QY 121 VVDCARDKNIPIRIGVNGSLEKDOEKYGEPTPOALLESAMRHVDHLDRNFDOFKVSV 180
DB 121 VVDCARDKNIPIRIGVNGSLEKDOEKYGEPTPOALLESAMRHVDHLDRNFDOFKVSV 180
QY 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARSAGVSAIGLILSEGIGDTLRVSL 240
DB 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARSAGVSAIGLILSEGIGDTLRVSL 240
QY 241 AADPVEEIKVGFIDILSLRIRANGINFACPTCSROEFVIGTVNMLEQRLEDITIPMDV 300
DB 241 AADPVEEIKVGFIDILSLRIRANGINFACPTCSROEFVIGTVNMLEQRLEDITIPMDV 300

```

```

QY 301 SIICGVNNGPGEALVSTLGVTGNGKSGLYEDGVRRDRLNMDIDLEARIKASQOLD 360
DB 301 SIICGVNNGPGEALVSTLGVTGNGKSGLYEDGVRRDRLNMDIDLEARIKASQOLD 360
QY 360 DEARRIDVQOYER 371
DB 360 DEARRIDVQOYER 371

```

## RESULT 6

```

F82283
gcpe protein VC0759 (imported) - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: F82283
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Baas, S.; Qin, H.; Dragol, I.; Sellers
L., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: AB2035; M01D:20406833; PMID:10952301
A:Accession: F82283
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-376 <HEI>
A:Cross-references: GB:AE004161; GB:AE003852; NID:9655200; PIDN:AAF93924.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0759
A:Map position: 1
C:Superfamily: gcpe protein

```

```

Query Match 86.9%; Score 1621.5; DB 2; Length 376;
Best Local Similarity 85.3%; Pred. No. 1.1e-104;
Matches 318; Conservative 33; Mismatches 21; Indels 1; Gaps 1;
QY 1 MHNOAPIORRKRSTRIYGVNPIGDCAPIAVOSMTNTRTDDVATVNOIKALERGVADIVR 60
DB 3 MOHESPIKRBPSTRIYGVNPIGDCAPIAVOSMTNTRTDDVATVNOIKALERGVADIVR 62

```

```

QY 61 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAVEYGVDCLRINPNGINEERIRIM 120
DB 63 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAVEYGVDCLRINPNGINEERIRIS 122
QY 121 VVDCARDKNIPIRIGVNGSLEKDOEKYGEPTPOALLESAMRHVDHLDRNFDOFKVSV 180
DB 123 VVDCARDKNIPIRIGVNGSLEKDOEKYGEPTPOALLESAMRHVDHLDRNFDOFKVSV 182
QY 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARSAGVSAIGLILSEGIGDTLRVSL 240
DB 183 KASDVFLAVESYRLAKQIDOPHLGITEAGARSAGVSAIGLILSEGIGDTLRVSL 242
QY 241 AADPVEEIKVGFIDILSLRIRANGINFACPTCSROEFVIGTVNMLEQRLEDITIPMDV 300
DB 243 AADPVEEIKVGFIDILSLRIRANGINFACPTCSROEFVIGTVNMLEQRLEDITIPMDV 302
QY 301 SIICGVNNGPGEALVSTLGVTGNGKSGLYEDGVRRDRLNMDIDLEARIKASQOLD 360
DB 303 SIICGVNNGPGEALVSTLGVTGNGKSGLYEDGVRRDRLNMDIDLEARIKASQOLD 362
QY 360 DEARRIDVQOYER 372
DB 363 DEARRIDVQOYER 375

```

## RESULT 7

```

H64063
gcpe protein - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jul-2000
C:Accession: H64063
R:Flieschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirschner, E.F.; Kerlavange

```











\_\_\_\_\_

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: August 12, 2003, 09:53:55 ; Search time 23 Seconds  
(without alignments)  
760,606 Million cell updates/sec

Title: US-09-921-992-78

Perfect score: 1866  
Sequence: 1 MHNQAFIQRRKSTRIVGNV.....RAKASQLEARRIDYQOVER 372

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1866	100.0	372	1	ISPG_ECOLI
2	1840	98.6	372	1	ISPG_SALTY
3	1831	98.1	372	1	ISPG_SALTY
4	1645.5	88.2	372	1	ISPG_VIBPA
5	1631.5	87.4	372	1	ISPG_VIBVU
6	1622.5	87.0	365	1	ISPG_PROST
7	1622.5	87.0	375	1	ISPG_YERPE
8	1621.5	86.9	376	1	ISPG_VIBCH
9	1603.5	85.9	368	1	ISPG_HAEIN
10	1599.5	85.7	367	1	ISPG_PASMU
11	1372.5	73.6	371	1	ISPG_PSEAE
12	1285.5	68.9	367	1	ISPG_BUCAP
13	1282	68.7	366	1	ISPG_WIGBR
14	1226.5	65.7	366	1	ISPG_WIGBR
15	1042.5	55.9	383	1	ISPG_CLOPR
16	906	48.6	349	1	ISPG_CLOPR
17	902.5	48.4	355	1	ISPG_THETN
18	887	47.5	349	1	ISPG_CLOAB
19	882.5	47.3	367	1	ISPG_BACHD
20	869.5	46.6	354	1	ISPG_FUSNN
21	868	46.5	368	1	ISPG_LISMO
22	859.5	46.1	377	1	ISPG_BACSU
23	848	44.4	357	1	ISPG_AQUAE
24	815.5	43.7	384	1	ISGI_STRCO
25	815	43.7	385	1	ISGI_STRCO
26	794	42.6	359	1	ISPG_HELPU
27	790	42.3	359	1	ISPG_HELPU
28	761	40.8	378	1	ISPG_CORGL
29	751.5	40.3	387	1	ISPG_MYCTU
30	712.5	38.2	392	1	ISPG_MYCLE
31	706	37.8	344	1	ISPG_THEMEA
32	690	37.0	357	1	ISPG_CAMUTE
33	618.5	33.1	403	1	ISPG_STNY3

## ALIGNMENTS

RESULT 1	ID	ISPG_ECOLI	STANDARD:	PRT:	372 AA.
AC	P27433	P76984	P76985		
DT	01-AUG-1992	(Rel. 23, Created)			
DT	01-AUG-1992	(Rel. 23, Last sequence update)			
DT	28-FEB-2003	(Rel. 41, Last annotation update)			
DE	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase (Gcpe protein)				
DE	(Protein E).				
GN	ISPG OR GCPE OR B2515 OR C3037 OR Z3778 OR ECS3377.				
OS	Escherichia coli, O6, and				
OS	Escherichia coli O157:H7.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
OC	Enterobacteriaceae; Escherichia.				
OX	NCBI_TaxID=562, 217992, 83334;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=K12;				
RC	MEDLINE=92394488; PubMed=1521767;				
RA	Baker J., Franklin D.B., Parker J.;				
RT	"Sequence and characterization of the gcpe gene of Escherichia coli.,"				
RL	FEMS Microbiol. Lett. 73:175-180(1992).				
RN	[2]				
RP	SEQUENCE FROM N.A., AND CHARACTERIZATION.				
RC	MEDLINE=97426617; PubMed=9278503;				
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,				
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,				
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,				
RA	Mau B., Shao Y.;				
RT	"The complete genome sequence of Escherichia coli K-12.,"				
RL	Science 277:123-145(1997).				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=K12;				
RC	MEDLINE=97349980; PubMed=9205837;				
RA	Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,				
RA	Itch T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,				
RA	Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,				
RA	Oshima T., Oyama S., Saito N., Stampel G., Satoh Y., Sivasubram S.,				
RA	Yamagata H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,				
RT	"Construction of a contiguous 874-kb sequence of the Escherichia coli				
RT	K-12 genome corresponding to 50.0-68.8 min on the linkage map and				
RT	analysis of its sequence features.,"				
RL	DNA Res. 4:91-113(1997).				
RN	[5]				

34 613.5 32.9 408 1 ISPG\_ANASP P58666 anabaena sp  
35 588 31.5 404 1 ISPG\_TREPA 083460 treponema p  
36 543.5 29.1 429 1 ISPG\_DELRA 091x60 delnoccocus  
37 527 28.2 421 1 ISPG\_NEIMA 09ju4 neisseria m  
38 527 28.2 421 1 ISPG\_NEIMB 09jz40 neisseria m  
39 512 27.4 417 1 ISPG\_RHITO 098f90 rhizobium l  
40 509 27.3 420 1 ISPG\_BRUME 08y17 bruceella m  
41 504.5 27.0 601 1 ISPG\_CHLMU 09px3 chlamydia m  
42 493 26.4 417 1 ISPG\_RHIME 092119 rhizobium m  
43 493 26.4 431 1 ISPG\_RALSO P58669 ralsotonia s  
44 492 26.4 416 1 ISPG\_AGRTS P58665 agrobacteri  
45 487 26.1 602 1 ISPG\_CHLTR 084060 chlamydia t

SEQUENCE FROM N.A.  
 RC STRAIN-06:H1 / CFT073 / ATCC 700928;  
 RC MEDLINE=22388234; PubMed=25471157;  
 RA Welch R.A., Burland V., Plunkett G., III, Redford P., Roesch P.,  
 RA Raeko D., Buckles E.L., Zhou S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Donnenberg K.S., Blattner F.R.;  
 RT "Extensive mosaic structures revealed by the complete genome sequence  
 of uropathogenic *Escherichia coli*.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).  
 [6]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-0157:H7 / EDL933 / ATCC 700927;  
 RC MEDLINE=21074935; PubMed=21206551;  
 RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lim A., Dimailanta E.T., Potamouis K.,  
 RA Apodaca J., Anantharaman P.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";  
 RL Nature 409:529-533 (2001).  
 [7]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-0157:H7 / RMD 050952;  
 RC MEDLINE=21156231; PubMed=21258796;  
 RA Hayashi T., Makino K., Onoishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Ikeda T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kihara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*  
 O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22 (2001).  
 [8]  
 RC PRELIMINARY SEQUENCE OF 343-372 FROM N.A.  
 RC MEDLINE=85261421; PubMed=2291272;  
 RA Freedman R., Gibson B., Donovan D., Biemann K., Eisenbeis S.J.,  
 RA Parker J., Schimmel P.;  
 RT "Primary structure of histidine-tRNA synthetase and characterization  
 of his transcripts.";  
 RL J. Biol. Chem. 260:10063-10068 (1985).  
 [9]  
 RC PATHWAY.  
 RC STRAIN-K12 / MC4100;  
 RC MEDLINE=21099853; PubMed=21163766;  
 RA Campos N., Rodriguez-Concepcion M., Seemann M., Rohmer M., Boronat A.;  
 RT "Identification of gcpE as a novel gene of the 2-C-methyl-D-erythritol  
 4-phosphate pathway for isoprenoid biosynthesis in *Escherichia coli*.";  
 RL FEBS Lett. 488:170-173 (2001).  
 [10]  
 RC PATHWAY.  
 RC MEDLINE=21172855; PubMed=21274098;  
 RA Altincicek B., Kollas A.-R., Sanderbrand S., Wiesner J., Hintz M.,  
 RA Beck F., Jomaa H.;  
 RT "GcpE is involved in the 2-C-methyl-D-erythritol 4-phosphate pathway  
 of isoprenoid biosynthesis in *Escherichia coli*.";  
 RL J. Bacteriol. 183:2411-2416 (2001).  
 [11]  
 CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate  
 (ME-2,4cpP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate.  
 CC -1- PATHWAY: Nonaevonate terpenoid biosynthesis pathway; sixth step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is Copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X64451; CAA45783.1;  
 DR EMBL: AY033515; AAK53460.2;  
 DR EMBL: AE000338; AAC75568.2;  
 DR

DR EMBL: D90880; BAA16402.1;  
 DR EMBL: D90881; BAA20919.1;  
 DR EMBL: AE016764; BAAB1487.1;  
 DR EMBL: AE005481; AAC57625.1;  
 DR EMBL: AP002561; BAB36800.1;  
 DR EMBL: M11843; -; NOT\_ANNOTATED\_CDS.  
 DR PIR: A91051; A91051.  
 DR PIR: E85895; E85895.  
 DR PIR: S23058; S23058.  
 DR ECoGene: EG10370; ISPG.  
 DR HAMAP: MF\_00159; -; 1.  
 DR InterPro: IPR004588; ISPG.  
 DR Pfam: PF04551; Gcpe; 1.  
 DR TIGRFAMs: TIGR00612; ISPG\_gcpe; 1.  
 KW Isoprene biosynthesis; Complete proteome.  
 SQ SEQUENCE 372 AA; 40683 MW; 3666FD83D5C04F9F CRC64;  
 Query Match 100.0%; Score 1866; DB 1; Length 372;  
 Best Local Similarity 100.0%; Pred. No. 1; 9e-121;  
 Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MHNAPIORRSTRIVGNVPIGDGAPIAVOSMTNTTDTVEATVNOIKALERVGADIVR 60  
 DB 1 MHNAPIORRSTRIVGNVPIGDGAPIAVOSMTNTTDTVEATVNOIKALERVGADIVR 60  
 QY 61 VSPPTMDAAEFKLKQOVNPLVADHFDYRIALKAAYGVDCIRINPNGINERIR 120  
 DB 61 VSPPTMDAAEFKLKQOVNPLVADHFDYRIALKAAYGVDCIRINPNGINERIR 120  
 QY 121 VVDCARKNITIRIGVNGNSLEKDOEKYGGEPDQALLESAMRHVDHDIRNPOFKSV 180  
 DB 121 VVDCARKNITIRIGVNGNSLEKDOEKYGGEPDQALLESAMRHVDHDIRNPOFKSV 180  
 QY 121 VVDCARKNITIRIGVNGNSLEKDOEKYGGEPDQALLESAMRHVDHDIRNPOFKSV 180  
 DB 121 VVDCARKNITIRIGVNGNSLEKDOEKYGGEPDQALLESAMRHVDHDIRNPOFKSV 180  
 QY 181 KASDFLAVESYRLAKIDQPLHLGITAGAGSAGVSAIGLLISEGIDTLRVSL 240  
 DB 181 KASDFLAVESYRLAKIDQPLHLGITAGAGSAGVSAIGLLISEGIDTLRVSL 240  
 QY 241 AADPEETKVGFDILKSRIRSGINFTACPTCSKQEDVIGYVALDQRLIEDITTPDV 300  
 DB 241 AADPEETKVGFDILKSRIRSGINFTACPTCSKQEDVIGYVALDQRLIEDITTPDV 300  
 QY 301 SIICGVNPGFALVSTLGTGNGKSGLYEDGVKRDLDNDMDIDQLEAIRAKASOLD 360  
 DB 301 SIICGVNPGFALVSTLGTGNGKSGLYEDGVKRDLDNDMDIDQLEAIRAKASOLD 360  
 QY 361 EARRIDVOQVER 372  
 DB 361 EARRIDVOQVER 372  
 RESULT 2  
 ISPG\_SALTY STANDARD; PRT; 372 AA.  
 AC P58671;  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.  
 GN ISPG OR Gcpe OR STM2523.  
 OS *Salmonella typhimurium*.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OC NCBI\_TaxID=602;  
 [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-LT2 / SGSC1412 / ATCC 700720;  
 RC MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium

```

RT L72.":
RL Nature 413:852-856(2001).
CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC (ME-2,4cpP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC (By similarity).
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE008814: AAL21417.1: -
DR SGENE: S6372727: ispg.
DR HAMAP: MF_00155: - 1.
DR InterPro: IPR006705: Gcpe.
DR InterPro: IPR004588: ispg.
DR Pfam: PF04551: Gcpe. 1.
DR TIGRfam: TIGR00612: ispg.gcpe. 1.
DR Isoprene biosynthesis: Complete proteome.
SQ SEQUENCE 372 AA: 40625 MW: 0486F8D38EC2AC9F CRC64:

Query Match 98.6%; Score 1840; DB 1; Length 372;
Best Local Similarity 98.7%; Pred. No. 1.1e-119;
Matches 367; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 MHNQAPIORRSTRIYGVNVPISGAPFAYVQSMTRTDTDEATVNOIKALERGVADIVR 60
DB 1 MHNQAPIORRSTRIYGVNVPISGAPFAYVQSMTRTDTDEATVNOIKALERGVADIVR 60
OY 61 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAVEYGVDCILRINPGINGNERIRM 120
DB 61 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAVEYGVDCILRINPGINGNERIRM 120
OY 61 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAVEYGVDCILRINPGINGNERIRM 120
DB 61 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAVEYGVDCILRINPGINGNERIRM 120
OY 121 VVDCARDKNIPRIYGVNAGSLEKDOEKGEPPTQALLSARHVDHLDRNFQFVSV 180
DB 121 VVDCARDKNIPRIYGVNAGSLEKDOEKGEPPTQALLSARHVDHLDRNFQFVSV 180
OY 121 VVDCARDKNIPRIYGVNAGSLEKDOEKGEPPTQALLSARHVDHLDRNFQFVSV 180
DB 121 VVDCARDKNIPRIYGVNAGSLEKDOEKGEPPTQALLSARHVDHLDRNFQFVSV 180
OY 181 KASDVFLAVESYRLAKOIDPPLHIGITTEAGARGAVKSAIGLLSSEGIGDILRVSL 240
DB 181 KASDVFLAVESYRLAKOIDPPLHIGITTEAGARGAVKSAIGLLSSEGIGDILRVSL 240
OY 241 AADPVEETKVGFDILKSLIRSRGINFTACPTCSROEDVIGTVNALEQRLEDITTPMDV 300
DB 241 AADPVEETKVGFDILKSLIRSRGINFTACPTCSROEDVIGTVNALEQRLEDITTPMDV 300
OY 301 SIIGCVNAGPEALVSTIGTGKNGKSGIYEDGVAKKRLDNDMDITDLEARIKASOLD 360
DB 301 SIIGCVNAGPEALVSTIGTGKNGKSGIYEDGVAKKRLDNDMDITDLEARIKASOLD 360
OY 361 EARRIDVOOVER 372
DB 361 EARRIDVOOVER 372
OY 361 EARRIDVOOVER 372
DB 361 EARRIDVOOVER 372

RESULT 3
ISPG_SALT1 STANDARD: PRT; 372 AA.
ID ISPG_SALT1 STANDARD: PRT; 372 AA.
AC P58670;
DR 28-FEB-2003 (Rel. 41, Created)
DR 28-FEB-2003 (Rel. 41, Last sequence update)
DR 15-SEP-2003 (Rel. 42, Last annotation update)
DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN ispg OR Gcpe OR STY2768 OR T0333.
OS Salmonella typh.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_Taxid=601;
[1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-CT18;
RX MEDLINE-21534947; PubMed-11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatina M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Cornerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagers K.,
RA Krogan A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Milledgehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typh CT18."
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Ty2 / ATCC 700931;
RX MEDLINE-22531367; PubMed-12644504;
RA Deng W., Lion S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typh strains Ty2
RT and CT18."
RL J. Bacteriol. 185:2330-2337(2003).
CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC (ME-2,4cpP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC (By similarity).
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL627275: CAD02726.1: -
DR EMBL: AE016835: AAO68056.1: -
DR HAMAP: MF_00155: - 1.
DR InterPro: IPR006705: Gcpe.
DR InterPro: IPR004588: ispg.
DR Pfam: PF04551: Gcpe. 1.
DR TIGRfam: TIGR00612: ispg.gcpe. 1.
DR Isoprene biosynthesis: Complete proteome.
SQ SEQUENCE 372 AA: 40626 MW: 3E3D5B13B8AA7916 CRC64:

Query Match 98.1%; Score 1831; DB 1; Length 372;
Best Local Similarity 98.1%; Pred. No. 4.8e-119;
Matches 365; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 MHNQAPIORRSTRIYGVNVPISGAPFAYVQSMTRTDTDEATVNOIKALERGVADIVR 60
DB 1 MHNQAPIORRSTRIYGVNVPISGAPFAYVQSMTRTDTDEATVNOIKALERGVADIVR 60
OY 61 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAVEYGVDCILRINPGINGNERIRM 120
DB 61 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAVEYGVDCILRINPGINGNERIRM 120
OY 121 VVDCARDKNIPRIYGVNAGSLEKDOEKGEPPTQALLSARHVDHLDRNFQFVSV 180
DB 121 VVDCARDKNIPRIYGVNAGSLEKDOEKGEPPTQALLSARHVDHLDRNFQFVSV 180
OY 121 VVDCARDKNIPRIYGVNAGSLEKDOEKGEPPTQALLSARHVDHLDRNFQFVSV 180
DB 121 VVDCARDKNIPRIYGVNAGSLEKDOEKGEPPTQALLSARHVDHLDRNFQFVSV 180
OY 181 KASDVFLAVESYRLAKOIDPPLHIGITTEAGARGAVKSAIGLLSSEGIGDILRVSL 240
DB 181 KASDVFLAVESYRLAKOIDPPLHIGITTEAGARGAVKSAIGLLSSEGIGDILRVSL 240
OY 241 AADPVEETKVGFDILKSLIRSRGINFTACPTCSROEDVIGTVNALEQRLEDITTPMDV 300
DB 241 AADPVEETKVGFDILKSLIRSRGINFTACPTCSROEDVIGTVNALEQRLEDITTPMDV 300
OY 301 SIIGCVNAGPEALVSTIGTGKNGKSGIYEDGVAKKRLDNDMDITDLEARIKASOLD 360
DB 301 SIIGCVNAGPEALVSTIGTGKNGKSGIYEDGVAKKRLDNDMDITDLEARIKASOLD 360

```



```

QY 301 SIIGCVVNGGGEALVSTLGTGKNGKSGLYEDGVR-KDRLDNDMDIDLEARIARASOL 359
DB 301 SIIGCVVNGGGEALVSTLGTAGSNKKSAPFIEDGKROKERFDNEDLVNQLAKIRAKARM 360
QY 360 DEARRIDVQ 368
DB 361 DESNRIDIK 369

RESULT 6
ISPG_PROST
ID ISPG_PROST STANDARD; PRT; 365 AA.
AC P7241;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN ISPG OR AACR.
OS Providencia stuartii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Providencia.
ON NCBI_TaxID=588;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PR50.
RX MEDLINE=97234638; PubMed=9079912;
RT Rather P.N., Solinsky K.A., Paradise M.R., Parojcic M.N.;
RT "Aaric, an essential gene involved in density-dependent regulation of
RT the 2',-N-acetyltransferase in Providencia stuartii."
RL J. Bacteriol. 179:2267-2273(1997).
CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC (ME-2,4CP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC (By similarity). Involved in density-dependent regulation of 2',-N-
CC acetyltransferase.
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DB HAMAP; MF_00159; -; 1.
DR InterPro; IPR004588; ISPG.
DR Pfam; PF04551; GCPE; 1.
DR TIGRFAMs; TIGR00612; ispg_gcpe; 1.
KW Isoprene biosynthesis.
SQ SEQUENCE 365 AA; 39840 MW; 4533147980480ECC CRC64;

Query Match 87.0%; Score 1622.5; DB 1; Length 365;
Best Local Similarity 88.4%; Pred. No. 1e-104;
Matches 320; Conservative 24; Mismatches 17; Indels 1; Gaps 1;

```

```

QY 241 AADVEEIKVGFEDILKSLRISRGINFACPTCSROEFVIGTVNMLEQRLIEDITPMY 300
DB 241 AADVEEIKVGFEDILKSLRISRGINFACPTCSROEFVIGTVNMLEQRLIEDITPMY 300
QY 301 SIIGCVVNGGGEALVSTLGTGKNGKSGLYEDGVR-KDRLDNDMDIDLEARIARASOL 359
DB 301 SIIGCVVNGGGEALVSTLGTAGAKTKSGYEDGVRKRRERDNDIIDLAKIRAKARM 360
QY 360 DE 361
DB 361 DE 362

RESULT 7
ISPG_YERPE
ID ISPG_YERPE STANDARD; PRT; 375 AA.
AC P58672;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN ISPG OR GCPE OR YPO2879 OR Y1353.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
ON NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Blovat Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sepaltia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Fellwell T., Hamlin N., Holtroyd S., Jagels R., Karlyshev A.V.,
RA Simmons M., Skellton J., Oyston P.C.F., Quail M., Rutherford K.,
RA "Genome sequence of Yersinia pestis, the causative agent of plague."
RL Nature 413:523-527(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM5 / Blovat Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Retherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Watson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM."
RL J. Bacteriol. 184:4601-4611(2002).
CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC (ME-2,4CP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC (By similarity).
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DB EMBL; AJ14154; CAC92130.1; -
DB EMBL; AE013738; AAM84926.1; -
DR PIR; AG0350; AG0350.
DR HAMAP; MF_00159; -; 1.
DR InterPro; IPR006705; GCPE.
DR InterPro; IPR004588; ISPG.
DR Pfam; PF04551; GCPE; 1.
DR TIGRFAMs; TIGR00612; ispg_gcpe; 1.

```

Isoprene biosynthesis: Complete proteome.  
 KW SEQUENCE 375 AA; 40797 MW; 860971872BIC0336 CRC64;  
 Query Match 87.0% Score 1622.5; DB 1; Length 375;  
 Best Local Similarity 86.8% Pred. No. 1.1e-104;  
 Matches 323; Conservative 19; Mismatches 29; Indels 1; Gaps 1;

QY 1 MHNQAPIORRKRSTRIYGVNVPIDGAPIAVQSMNTFTTVDVATVQIKLERYGADIVR 60  
 1 MHNSPIIRKRSTRIYGVNVPIDGAPIAVQSMNTFTTVDVATVQIKLERYGADIVR 60  
 DB 1 MHNSPIIRKRSTRIYGVNVPIDGAPIAVQSMNTFTTVDVATVQIKLERYGADIVR 60  
 QY 61 VSVPTMDAAAFKLIKQVSVPLVADHFDPRIALKVAEYGVDCRLNPGNIGNERIRM 120  
 61 VSVPTMDAAAFKLIKQVSVPLVADHFDPRIALKVAEYGVDCRLNPGNIGNERIRM 120  
 DB 61 VSVPTMDAAAFKLIKQVSVPLVADHFDPRIALKVAEYGVDCRLNPGNIGNERIRM 120  
 QY 121 VVDCARDKNIPIRIGVAGSLKEDLOEKYGEPTPOALLSARHVDLRLNFDQKVS 180  
 121 VVDCARDKNIPIRIGVAGSLKEDLOEKYGEPTPOALLSARHVDLRLNFDQKVS 180  
 DB 121 VVDCARDKNIPIRIGVAGSLKEDLOEKYGEPTPOALLSARHVDLRLNFDQKVS 180  
 QY 181 KASVFLAVESYRLAQTIDQPLHLGITEAGGASGAVKSAIGILLSEIGDTLRVSL 240  
 181 KASVFLAVESYRLAQTIDQPLHLGITEAGGASGAVKSAIGILLSEIGDTLRVSL 240  
 DB 181 KASVFLAVESYRLAQTIDQPLHLGITEAGGASGAVKSAIGILLSEIGDTLRVSL 240  
 QY 241 AADVEETKVGFDILKSLRIRSGINFIACPTCSROEFDVIGYVNALEQRLIEDIITPM 300  
 241 AADVEETKVGFDILKSLRIRSGINFIACPTCSROEFDVIGYVNALEQRLIEDIITPM 300  
 DB 241 AADVEETKVGFDILKSLRIRSGINFIACPTCSROEFDVIGYVNALEQRLIEDIITPM 300  
 QY 301 SIICVYVNGPEALVSTLGTGKNGKSGLYEDGYR-KDRLNDNDMDIOLKRIKAKASOL 359  
 301 SIICVYVNGPEALVSTLGTGKNGKSGLYEDGYR-KDRLNDNDMDIOLKRIKAKASOL 359  
 DB 301 SIICVYVNGPEALVSTLGTGKNGKSGLYEDGYR-KDRLNDNDMDIOLKRIKAKASOL 359  
 QY 360 DEARRIDVOQVE 371  
 360 DEARRIDVOQVE 371  
 DB 361 DANNRIVNQLD 372  
 361 DANNRIVNQLD 372

RESULT 8  
 ISPG\_VIBCH STANDARD; PRT: 376 AA.  
 ID ISPG\_VIBCH STANDARD; PRT: 376 AA.  
 AC 09KTX1:  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.  
 GN ISPG OR VC0759.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=El Tor N16961 / Serotype O1;  
 RX MEDLINE=20406833; PubMed=30952301;  
 RA Heidelberg J.F., Eissen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.R., Peterson J.D., Unayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 RA McDonald L., Utterback T., Fieschmann T.D., Mierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 RT cholerae";  
 RL Nature 406:477-483(2000).  
 CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate  
 CC (ME-2,4CP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate  
 CC (BY similarity).  
 CC -1- PATWAY: Nonmevalonate terpenoid biosynthesis pathway, sixth step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE004161; AAF93924.1; -.  
 CC PIR: F82283; F82283.  
 CC TIGR: VC0759; -.  
 CC HAMAP: MF\_00159; -; 1.  
 CC InterPro: IPR004588; Ispg.  
 CC Pfam: PF04551; Gcpe; 1.  
 CC TIGRFAMs: TIGR00612; Ispg-gcpe; 1.  
 KW Isoprene biosynthesis: Complete proteome.  
 KW SEQUENCE 376 AA; 40862 MW; 34E4144CB7ADAAA CRC64;  
 Query Match 86.9% Score 1621.5; DB 1; Length 376;  
 Best Local Similarity 85.3% Pred. No. 1.3e-104;  
 Matches 318; Conservative 33; Mismatches 21; Indels 1; Gaps 1;

QY 1 MHNQAPIORRKRSTRIYGVNVPIDGAPIAVQSMNTFTTVDVATVQIKLERYGADIVR 60  
 1 MHNSPIIRKRSTRIYGVNVPIDGAPIAVQSMNTFTTVDVATVQIKLERYGADIVR 62  
 DB 3 MHNSPIIRKRSTRIYGVNVPIDGAPIAVQSMNTFTTVDVATVQIKLERYGADIVR 62  
 QY 61 VSVPTMDAAAFKLIKQVSVPLVADHFDPRIALKVAEYGVDCRLNPGNIGNERIRM 120  
 61 VSVPTMDAAAFKLIKQVSVPLVADHFDPRIALKVAEYGVDCRLNPGNIGNERIRM 122  
 DB 63 VSVPTMDAAAFKLIKQVSVPLVADHFDPRIALKVAEYGVDCRLNPGNIGNERIRM 122  
 QY 121 VVDCARDKNIPIRIGVAGSLKEDLOEKYGEPTPOALLSARHVDLRLNFDQKVS 180  
 121 VVDCARDKNIPIRIGVAGSLKEDLOEKYGEPTPOALLSARHVDLRLNFDQKVS 182  
 DB 123 VVDCARDKNIPIRIGVAGSLKEDLOEKYGEPTPOALLSARHVDLRLNFDQKVS 182  
 QY 181 KASVFLAVESYRLAQTIDQPLHLGITEAGGASGAVKSAIGILLSEIGDTLRVSL 240  
 181 KASVFLAVESYRLAQTIDQPLHLGITEAGGASGAVKSAIGILLSEIGDTLRVSL 242  
 DB 183 KASVFLAVESYRLAQTIDQPLHLGITEAGGASGAVKSAIGILLSEIGDTLRVSL 242  
 QY 241 AADVEETKVGFDILKSLRIRSGINFIACPTCSROEFDVIGYVNALEQRLIEDIITPM 300  
 241 AADVEETKVGFDILKSLRIRSGINFIACPTCSROEFDVIGYVNALEQRLIEDIITPM 302  
 DB 243 AADVEETKVGFDILKSLRIRSGINFIACPTCSROEFDVIGYVNALEQRLIEDIITPM 302  
 QY 301 SIICVYVNGPEALVSTLGTGKNGKSGLYEDGYR-KDRLNDNDMDIOLKRIKAKASOL 359  
 301 SIICVYVNGPEALVSTLGTGKNGKSGLYEDGYR-KDRLNDNDMDIOLKRIKAKASOL 362  
 DB 303 SIICVYVNGPEALVSTLGTGKNGKSGLYEDGYR-KDRLNDNDMDIOLKRIKAKASOL 362  
 QY 360 DEARRIDVOQVE 372  
 360 DEARRIDVOQVE 372  
 DB 363 DEKNRIDIKHVEQ 375  
 363 DEKNRIDIKHVEQ 375

RESULT 9  
 ISPG\_HAEIN STANDARD; PRT: 368 AA.  
 ID ISPG\_HAEIN STANDARD; PRT: 368 AA.  
 AC P44667:  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.  
 GN ISPG OR GCPE OR HI0368.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Rd / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fieschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Keriavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Goeyne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Utterback T.R., Hanna M.C., Fuhrmann J.L., Geoghegan N.S.M.,  
 RA Fine L.D., Fritchman J.L., Nguyen D.T., Sauder D.M., Brandon R.C.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;



```

RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=pm70;
RA      MEDLINE=21145866; PubMed=11248100;
RA      May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT      "Complete genomic sequence of Pasteurella multocida pm70."
RL      Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC      -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC      (ME-2,4CpP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC      (by similarity).
CC      -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; slxth step.
CC      -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC      -----
CC      This Swiss-Prot entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on
CC      use by non profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC      or send an email to license@isb-slb.ch).
CC
DR      EMBL; AE006237; AAK0094.1; -
DR      HAMAP; MF_00159; -; 1
DR      InterPro; IPR004588; ISPG.
DR      Pfam; PF04551; IGSP; 1.
DR      TIGRFAMs; TIGR00612; ISPG_gcPE; 1.
DR      Isoprene biosynthesis; Complete proteome.
SQ      SEQUENCE 367 AA; 39669 MW; 1918C5702642AEE8 CRC64;

Query Match      85.7%; Score 1599.5; DB 1; Length 367;
Best Local Similarity 88.9%; Pred. NO. 4e-103;
Matches 320; Conservative 18; Mismatches 21; Indels 1; Gaps 1,

QY      7 IQRKSTIYGVNVPVIGDAPLAVGSMNTRTDVEATVNOIKLERYGADIVRSYPTM 66
DB      7 IIRKSTIYGVNVPVIGDAPLAVGSMNTRTDVEATVNOIKLERYGADIVRSYPTM 66
QY      7 IKRRESTIYGVNVPVIGDAPLAVGSMNTRTDVEATVNOIKLERYGADIVRSYPTM 66
DB      7 IKRRESTIYGVNVPVIGDAPLAVGSMNTRTDVEATVNOIKLERYGADIVRSYPTM 66
QY      67 DAAPAFKIKQOVNPLVADHDFEDRLALKVAEYGVDCRLINPNGINIGERIIRNVDCAR 126
DB      67 DAAPAFKIKQOVNPLVADHDFEDRLALKVAEYGVDCRLINPNGINIGERIIRNVDCAR 126
QY      127 DKNPIRIRGVNAGSLERKDLQEKYGEPPVQALLGSMRNVHDLRLNPFQFVYSKASVF 186
DB      127 DKNPIRIRGVNAGSLERKDLQEKYGEPPVQALLGSMRNVHDLRLNPFQFVYSKASVF 186
QY      127 DNNPIRIRGVNAGSLERKDLQEKYGEPPVQALLGSMRNVHDLRLNPFQFVYSKASVF 186
DB      127 DNNPIRIRGVNAGSLERKDLQEKYGEPPVQALLGSMRNVHDLRLNPFQFVYSKASVF 186
QY      187 LAVESYRLAKQIDQPLHIGITTEAGGARGAVKSAIGLLSGISGDTLRFVSLAADPYE 246
DB      187 LAVESYRLAKQIDQPLHIGITTEAGGARGAVKSAIGLLSGISGDTLRFVSLAADPYE 246
QY      187 LAVESYRLAKQIDQPLHIGITTEAGGARGAVKSAIGLLSGISGDTLRFVSLAADPYE 246
DB      187 LAVESYRLAKQIDQPLHIGITTEAGGARGAVKSAIGLLSGISGDTLRFVSLAADPYE 246
QY      247 EIKVGPILMSLRIRSGINFIACPTCSROEFDVIQTVNALEQRLEDITTPMDYSITGV 306
DB      247 EIKVGPILMSLRIRSGINFIACPTCSROEFDVIQTVNALEQRLEDITTPMDYSITGV 306
QY      247 EIKVGPILMSLRIRSGINFIACPTCSROEFDVIQTVNALEQRLEDITTPMDYSITGV 306
DB      247 EIKVGPILMSLRIRSGINFIACPTCSROEFDVIQTVNALEQRLEDITTPMDYSITGV 306
QY      307 VNGGEALVSLVSLGYGVGNKKSGLEVDEQVR-KDRLDNDMDIDQLEARIARASQDEARRI 365
DB      307 VNGGEALVSLVSLGYGVGNKKSGLEVDEQVR-KDRLDNDMDIDQLEARIARASQDEARRI 365
QY      307 VNGGEALVSLVSLGYGVGNKKSGLEVDEQVR-KDRLDNDMDIDQLEARIARASQDEARRI 365
DB      307 VNGGEALVSLVSLGYGVGNKKSGLEVDEQVR-KDRLDNDMDIDQLEARIARASQDEARRI 365

RESULT 11
ISPG_PSEAE
ID ID ISPG_PSEAE STANDARD; PRT; 371 AA.
AC O9HX14;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
ISPG OR PA3803.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
ON NCBI_TaxID=287;
XP [1]
SEQUENCE FROM N.A.

```

RC STRAIN-ATCC 15692 / PAOI;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.B., Hutnagan W.O., Kowalik D.J., Lagrou M.,  
 RA Gader R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardiz K., Lim K.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RA "Complete genome sequence of *Pseudomonas aeruginosa* PAOI, an  
 opportunistic pathogen.";  
 RT Nature 406:959-964(2000).  
 CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate  
 CC (ME-2,4cPP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate  
 CC (By similarity).  
 CC -1- PATHWAY: Nonnevalonate (terpenoid) biosynthesis pathway; sixth step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE004798; AAC07190.1;  
 CC PIR: F83171; F83171.  
 CC HAMAP: MF\_00159; -; 1.  
 CC InterPro: IPR004588; ISPG.  
 CC Pfam: PF04551; GCPE; 1.  
 CC TIGRFAMs: TIGR00612; ISPG\_GCPG; 1.  
 CC Isoprene biosynthesis: Complete proteome.  
 KW SEQUENCE 371 AA; 40056 MW; F33FC1PCE017117 CRC64;  
 SQ  
 Query Match 73.63% Score 1372.5; DB 1; Length 371;  
 Best Local Similarity 73.83% Pred. No. 1.7e-87;  
 Matches 267; Conservative 46; Mismatches 48; Indels 1; Gaps 1;  
 QY 1 MNHNPIDRRKSTRIVYGNVPIGDCAPIVQSMNTRTTDEATVNOIKALERGVADIVR 60  
 DB 3 IHSASITIRKRSKRIKVGWVNPVIGDAPISVQSMNTTETCDVAATVQIFRLEADAGADIVR 62  
 QY 61 VSVPTMDAAEAFKLKIQVNVPLVADHEDYRIALKAVEYGVDCLRINPGNIGNEERIRM 120  
 DB 63 VSVPTMDAAEAFKLKIQVNVPLVADHEDYRIALKAVEYGVDCLRINPGNIGNEERIRYA 122  
 QY 121 VVDCARDKNIPRIGVNGSLKEDLOEKYGEPTPOLLESAMRHVDHLRLNFDQFVSV 180  
 DB 123 VVDAARERNIPRIGVNGSLKEDLOEKYGEPTPOLLESAMRHVDHLRLNFDQFVSV 182  
 QY 181 KASDVLAVESYRLAKIQIDPLHLGITTEAGARGAVKSAIGLLSEGIGDTLRVSL 240  
 DB 183 KASDVLAVESYRLAKIQIDPLHLGITTEAGARGAVKSAIGLLSEGIGDTLRVSL 242  
 QY 241 AADPVEIKVGDILKSLRIRSGINFIACPTCSQEPEDVIGVNALRQRLIEDIITPMV 300  
 DB 243 AADPVEIKVGDILKSLRIRSGINFIACPTCSQEPEDVIGVNALRQRLIEDIITPMV 302  
 QY 301 SIIGCVVNGEALVSTLGVGTGANKSGLYEDGVDRDLNDNDMDIDOLARIRAKASQL 360  
 DB 303 SIIGCVVNGEALVSTLGVGTGANKSGLYEDGVDRDLNDNDMDIDOLARIRAKASQL 361  
 QY 361 EA 362  
 DB 362 EA 363  
 RESULT 12  
 ISPG\_BUCAP STANDARD; PRT; 367 AA.  
 AC 08K9P4;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.  
 GN ISPG OR GCPE OR B0SG276.  
 OS Buchnera aphidicola (subsp. *Schizaphis graminum*).  
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC Enterobacteriaceae; Buchnera.  
 OX NCBI\_TaxID=98794;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22084549; PubMed=12089438;  
 RA Tamas I., Klasson L., Canhaeck B., Naeslund A.K., Eriksson A.-S.,  
 RA Wernegren J.J., Sandstrom J.P., Moran N.A., Andersson S.G.E.;  
 RT "50 million years of genomic stasis in endosymbiotic bacteria.";  
 RT Science 296:2376-2379(2002).  
 CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate  
 CC (ME-2,4cPP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate  
 CC (By similarity).  
 CC -1- PATHWAY: Nonnevalonate (terpenoid) biosynthesis pathway; sixth step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE014104; AAM67834.1;  
 CC HAMAP: MF\_00159; -; 1.  
 CC InterPro: IPR006705; GCPE.  
 CC InterPro: IPR004588; ISPG.  
 CC Pfam: PF04551; GCPE; 1.  
 CC TIGRFAMs: TIGR00612; ISPG\_GCPG; 1.  
 CC Isoprene biosynthesis: Complete proteome.  
 KW SEQUENCE 367 AA; 40440 MW; 4CF6A0FAF017661C CRC64;  
 SQ  
 Query Match 68.9% Score 1285.5; DB 1; Length 367;  
 Best Local Similarity 66.7% Pred. No. 1.6e-81;  
 Matches 244; Conservative 64; Mismatches 57; Indels 1; Gaps 1;  
 QY 1 MNHNPIDRRKSTRIVYGNVPIGDCAPIVQSMNTRTTDEATVNOIKALERGVADIVR 60  
 DB 1 MNKYNIRKRSKRIKVGWVNPVIGDAPISVQSMNTTETCDVAATVQIFRLEADAGADIVR 62  
 QY 61 VSVPTMDAAEAFKLKIQVNVPLVADHEDYRIALKAVEYGVDCLRINPGNIGNEERIRM 120  
 DB 61 VSVPTMDAAEAFKLKIQVNVPLVADHEDYRIALKAVEYGVDCLRINPGNIGNEERIRL 120  
 QY 121 VVDCARDKNIPRIGVNGSLKEDLOEKYGEPTPOLLESAMRHVDHLRLNFDQFVSV 180  
 DB 121 VVDCARDKNIPRIGVNGSLKEDLOEKYGEPTPOLLESAMRHVDHLRLNFDQFVSV 180  
 QY 181 KASDVLAVESYRLAKIQIDPLHLGITTEAGARGAVKSAIGLLSEGIGDTLRVSL 240  
 DB 181 KASDVLAVESYRLAKIQIDPLHLGITTEAGARGAVKSAIGLLSEGIGDTLRVSL 240  
 QY 241 AADPVEIKVGDILKSLRIRSGINFIACPTCSQEPEDVIGVNALRQRLIEDIITPMV 300  
 DB 241 AADPVEIKVGDILKSLRIRSGINFIACPTCSQEPEDVIGVNALRQRLIEDIITPMV 300  
 QY 301 SIIGCVVNGEALVSTLGVGTGANKSGLYEDGVDRDLNDNDMDIDOLARIRAKASQL 359  
 DB 301 SIIGCVVNGEALVSTLGVGTGANKSGLYEDGVDRDLNDNDMDIDOLARIRAKASQL 360  
 QY 360 DEARRI 365  
 DB 361 DKIKKI 366  
 RESULT 13  
 ISPG\_BUCAI STANDARD; PRT; 368 AA.  
 ID ISPG\_BUCAI

AC PS7374;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.  
 GN ISPG OR GCPE OR BU287.  
 OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum  
 symbiotic bacterium).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Buchnera.  
 OX NCBI\_TaxID:118099;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Tokyo 1998;  
 RX MEDLINE-20445173; PubMed-1093077;  
 RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;  
 RT "Genome sequence of the endocellular bacterial symbiont of aphids  
 Buchnera sp. APS.";  
 RT Nature 407:81-86(2000).  
 CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate  
 (ME-2,4cpP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate  
 (By similarity).  
 CC -1- PATHWAY: Nomenclature terpenoid biosynthesis pathway; sixth step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AP001118; BAB12997.1; -  
 DR HSSP; P49058; IEPP.  
 DR HAAAP; MF\_00159; -; 1.  
 DR InterPro; IPR004588; ISPG.  
 DR Pfam; PF04551; GCPE; 1.  
 DR TIGRFAMs; TIGR00612; ISPG\_gcpe; 1.  
 DR Isoprene biosynthesis; Complete proteome.  
 KW SEQUENCE 368 AA; 40869 MW; 8BC32F495954C1DE CRC64;  
 SQ  
 Query Match 68.7%; Score 1282; DB 1; Length 368;  
 Best Local Similarity 66.1%; Pred. No. 2,9e-81;  
 Matches 242; Conservative 65; Mismatches 57; Indels 2; Gaps 2;  
 QY 1 MHNQAP-IQRKSTRIVGNVPIGDCAPIAVOSMTNTTDEATVNOIKALERVGADIV 59  
 DB 1 MNKKKTIINRRKSDRIYGVKVAIGNNAPISVOSMTNTTSETINQILOKVGVDIV 60  
 QY 60 RVSVPMDAAEAEKLIKQOVNPLVADIHEDYRIALKAVEGVDCIRINPGINGEIR 119  
 DB 61 RISIPMLKAESKEKKTQTNPLADIDHFDYRIALQAIKYGADCLIRINPGINGRVRS 120  
 QY 120 MYVDCARKNIPRIGVANGSLEKDLQEKYGEPTPOLLESAMRHVDHLRINFGOFKVS 179  
 DB 121 EISYAKDENIPRIGVANGSLEKDLKIKYKIPTPALVESAMRHIEYDALNFOFKVS 180  
 QY 180 VKASDFELAVESRIALQIDPRLHGTETAGAGSAGVSAIGCLISBISGDTLRVS 239  
 DB 181 VKASDFELAVESRIALQIDPRLHGTETAGAGSAGVSAIGCLISBISGDTLRVS 240  
 QY 240 LAADPEETKIVGDFILKSLIRSGINFIACPTCSROEFDVGTYNALAEORLEDIITPMD 299  
 DB 241 LAADPEETKIVGDFILKSLIRSGINFIACPTCSROEFDVGTYNALAEORLEDIITPMD 300  
 QY 300 VSIIGCVVNGIGESKIALTGLAGSHKSAFEDGVROKEIKINEIIEKMEIKIRKIDK 358  
 DB 301 VSIIGCVVNGIGESKIALTGLAGSHKSAFEDGVROKEIKINEIIEKMEIKIRKIDK 360  
 QY 359 LDEARR 364  
 DB 361 LNNKK 366

RESULT 14  
 ID ISPG\_WIGBR STANDARD; PRT; 366 AA.  
 AC 08D1Y3;  
 DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.  
 GN ISPG OR GCPE OR WIGBR5730.  
 OS Wigglesworthia glossinidia brevipalpis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Wigglesworthia.  
 OX NCBI\_TaxID:36870;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE-22297718; PubMed-12219091;  
 RX Akman L., Yamashita A., Watanabe H., Oshima K., Shiba T., Hattori M.,  
 RA Aksoy S.;  
 RT "Genome sequence of the endocellular obligate symbiont of tsetse  
 flies, Wigglesworthia glossinidia.";  
 RT Nat. Genet. 32:402-407(2002).  
 CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate  
 (ME-2,4cpP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate  
 (By similarity).  
 CC -1- PATHWAY: Nomenclature terpenoid biosynthesis pathway; sixth step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB063522; BAC24719.1; -  
 DR HAAAP; MF\_00159; -; 1.  
 DR InterPro; IPR006705; GCPE.  
 DR InterPro; IPR004588; ISPG.  
 DR Pfam; PF04551; GCPE; 1.  
 DR TIGRFAMs; TIGR00612; ISPG\_gcpe; 1.  
 DR Isoprene biosynthesis; Complete proteome.  
 KW SEQUENCE 366 AA; 40407 MW; 159E01698F08DED CRC64;  
 SQ  
 Query Match 65.7%; Score 1226.5; DB 1; Length 366;  
 Best Local Similarity 65.9%; Pred. No. 1.9e-77;  
 Matches 234; Conservative 62; Mismatches 58; Indels 1; Gaps 1;  
 QY 7 IQRRKSTRIVGNVPIGDCAPIAVOSMTNTTDEATVNOIKALERVGADIVSVPTM 66  
 DB 7 IIRRSKRIYKNNPIGSDSPISVOSMTNTTINISQINKLQARADIVKVSIPTL 66  
 QY 67 DAAEFKLIKQOVNPLVADIHEDYRIALKAVEGVDCIRINPGINGEIRIRVYDCAR 126  
 DB 67 EAAESFKIKRNVSPIVADIHEDYRIALKAEEGADCLIRINPGINCKLRIRISVSTAK 126  
 QY 127 DNKIPRIGVANGSLEKDLQEKYGEPTPOLLESAMRHVDHLRINFGOFKVSADVF 186  
 DB 127 EKKIPRIGVANGSLEKDLQEKYGEPTPOLLESAMRHVDHLRINFGOFKVSADVF 186  
 QY 187 LAVESYRLAKQIDPRLHGTETAGAGSAGVSAIGCLISBISGDTLRVSLADPVE 246  
 DB 187 TVQVSYRLAKQIDPRLHGTETAGAGSAGVSAIGCLISBISGDTLRVSLADPVE 246  
 QY 247 EIKYGFILKSLIRSGINFIACPTCSROEFDVGTYNALAEORLEDIITPMDVSIIGCV 306  
 DB 247 EVKYGESILSLNRRKGINFIACPTCSROEFDVGTYNALAEORLEDIITPMDVSIIGCV 306  
 QY 307 VNGPEALVSTLGTGNGKSGLYEDGV-RKDLRNNDMIDQLEARRAKASQD 360  
 DB 307 VNGPEALVSTLGTGNGKSGLYEDGV-RKDLRNNDMIDQLEARRAKASQD 361

```

RESULT 15
ISPG.CAUCR STANDARD; PRT: 383 AA.
ID ISPG.CAUCR
AC 09A9W0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN ISPG OR CC0851.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 19089 / CB15;
RC MEDLINE=21173698; Pubmed=38259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterberg T., Tran K., Moll A., Yamathayan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RT Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC (ME-2,4cpP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC (By similarity).
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AE005761; AAK22836.10.
DR DR PIR: H87354; H87354.
DR TIGR: CC0851.
DR HAMAP: MF_00159; -.
DR InterPro: IPR004588; ISPG.
DR Pfam: PF04551; GCPE; 1.
DR TIGRPFAMs: TIGR00612; ISPG_gcpe; 1.
DR Isoprene biosynthesis; Complete proteome.
SQ SEQUENCE 383 AA; 40788 MW; 8468E8400E338194 CRC64;

Query Match 55 936 Score 1042.5; DB 1; Length 383;
Best Local Similarity 57.63; Pred. No. 9.2e-65;
Matches 208; Conservative 57; Mismatches 95; Indels 1; Gaps 1;

OY 7 IORRSTRIYGVNPIIGCAPIAVOSMTNRTTVDVATVNOIKALERGVADIVRSVPTM 66
DB 14 ITRRQSRKIRYGVSEVGGADAPISVQSMVTNLTSDAATLEQIROLFEAGADIVRSVCPDV 73
OY 67 DAAEAFKLIKQVNVPEVDIHFYRIALKAVEYGVDCIRINPNTGNENRIRRVYDCAR 126
DB 74 ESTAAFKTIRAEKVEVDIHFHYRGJIEAQAAGAACLRINPNTGNISPDVRVDVIOAAR 133
OY 127 DKNIPIRIGVNASLKEEDLOEKYGEPTPOALLSARHVDHLRLNFDQFKSVKASDVF 186
DB 134 DHCCSMKIGVNASLKEEDLOEKYGEPTPOALLSARHVDHLRLNFDQFKSVKASDVF 193
OY 187 LAVESYRLAKQIDQPEHLGITEAGARGAVKSAIGLILSEGIIDTLRVSADPVE 246
DB 194 MYVAAYYQLAIDCPEHLGITEAGARGAVKSAIGLILSEGIIDTLRVSADPVE 253
OY 247 EIKVGEDILKSLIRSGINFIACPTCSROEFDVIGTVNALBQRLEDIITPMDSIIGCV 306

```

```

DB 254 EIKVGEDILKSLIRSGINFIACPTCSROEFDVIGTVNALBQRLEDIITPMDSIIGCV 313
OY 307 VNGPGEALVSTLGVTGNNKSGL-YEDGYRKRLDNNMDIDLEARIKAKASQLEEARRI 365
DB 314 VNGPGEALMTDIGFTGGAGAGMYWAGKPDHKQSNEMIDHIVLVEKKAIEIOAKAQ 373
OY 366 D 366
DB 374 D 374

```

Search completed: August 12, 2003, 09:54:29  
Job time : 25 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 14, 2003, 21:54:20 ; Search time 2622 seconds

(without alignments)  
3448.232 Million cell updates/sec

Title: US-09-921-992-78  
Perfect score: 1866  
Sequence: 1 MHNOAPIQRKSTRIVGVN.....RAKASQLDEARRIDYQOVER 372

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+pn.model -DEV=zh  
-Q/cgn2.1/USPTO.spool/US09921992/runat\_12082003.094912.17379/app\_query.fasta\_1.519  
-DB=EST -QFMT=fastap -SUFFIX=st -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdt LIST=45  
-DOCALIGN=200 -THR\_SCORE=dot -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext HEADSITE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09921992.cgn2.1.2810.crunat\_12082003.094912.17379 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEUDERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST\*\*  
1: em\_estda:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estmu:\*  
5: em\_estrov:\*  
6: em\_estrpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vit:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_tod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vr1:\*  
28: gb\_gss1:\*

29: gb\_gss2:\*  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	820.5	44.0	1323	29	B2575759	B2575759 msh2_4614
2	379	20.3	828	14	CB669733	CB669733 OSJNE02H
3	348.5	18.7	853	14	CB619369	CB619369 OSITE030
4	326.5	17.5	1442	11	AY104363	AY104363 Zea mays
5	321.5	17.2	1268	29	B2569685	B2569685 pac2-164
6	315.5	16.9	579	12	BI920888	BI920888 EST540823
7	313	16.8	720	28	BH235010	BH235010 PH.05.X.S
8	307.5	16.5	658	28	A2570993	A2570993 281PVD07
9	276.5	14.8	852	14	CB628479	CB628479 OSITEB04B
10	271	14.5	602	14	CA022320	CA022320 H242M11r
11	269.5	14.4	610	12	BJ481080	BJ481080 BJ481080
12	265	14.2	706	10	BG591263	BG591263 EST499105
13	264	14.1	489	6	AW042702	AW042702 Porphyra
14	262	14.0	601	9	AW042702	AW042702 ST23B04.P
15	261	14.0	603	12	BJ551332	BJ551332 BJ551332
16	260	13.9	558	9	AV917069	AV917069 AV917069
17	260	13.9	608	12	BJ465443	BJ465443 BJ465443
18	260	13.9	616	9	AV934187	AV934187 AV934187
19	259	13.9	484	13	BO080994	BO080994 san12909
20	259	13.9	651	9	AU252386	AU252386 AU252386
21	257	13.8	640	13	BO091269	BO091269 st70g12.Y
22	252	13.5	579	28	A2523878	A2523878 224PBE02
23	250.5	13.4	351	13	BU004160	BU004160 OG637N12
24	244	13.1	465	9	A182869	A182869 EST242192
25	236	12.6	576	13	BQ975044	BQ975044 QH118D05
26	232.5	12.5	545	9	AV626844	AV626844 AV626844
27	229	12.3	532	10	BE924238	BE924238 EST428007
28	229	12.3	628	13	BU048059	BU048059 PR_1EP003
29	229	12.3	680	12	BB890681	BB890681 EST516532
30	229	12.3	700	12	BI933447	BI933447 EST516532
31	229	12.3	728	10	BG525868	BG525868 53-47-1-2
32	228	12.2	423	13	BO862800	BO862800 OGC22B19
33	225	12.1	792	14	CB620208	CB620208 OSITEA05D
34	225	12.1	821	14	CB291771	CB291771 UCRCS01.0
35	225	12.1	821	14	CB628414	CB628414 OSITEB03P
36	225	12.1	1127	29	B2577824	B2577824 msh2_5589
37	224	12.0	687	14	CA781720	CA781720 046F07AF
38	223.5	11.9	413	13	BM059608	BM059608 K501002EO
39	222.5	11.9	544	13	BO591926	BO591926 E012583-0
40	222.5	11.9	731	10	BG124857	BG124857 EST470503
41	222	11.9	640	9	AW738440	AW738440 EST339867
42	218.5	11.7	770	14	CB662235	CB662235 OSJNE030B
43	216	11.6	448	9	AV644003	AV644003 AV644003
44	216	11.6	482	9	AV642898	AV642898 AV642898
45	216	11.6	502	9	AV626792	AV626792 AV626792

## ALIGNMENTS

RESULT 1  
LOCUS B2575759 1323 bp DNA linear GSS 17-DEC-2002  
DEFINITION msh2\_4614.x1 msh Pseudomonas aeruginosa genomic clone msh2\_4614,  
B2575759 genomic survey sequence.  
ACCESSION B2575759  
VERSION B2575759.1 GI:27210820  
KEYWORDS GSS.  
SOURCE Pseudomonas aeruginosa  
ORGANISM Pseudomonas aeruginosa  
REFERENCE 1 (bases 1 to 1323)  
Bacteria: Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.

**AUTHORS** Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M., Burns, J.L., Kaul, R. and Olsen, M.V.  
**TITLE** Whole-Genome-Sequence variation among multiple isolates of *Pseudomonas aeruginosa* library  
**JOURNAL** J. Bacteriol. (2002) in press  
**COMMENT** Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: craymond@u.washington.edu

**FEATURES**  
 source  
 Location/Qualifiers  
 1..1323  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="MSH"  
 /db\_xref="taxon:287"  
 /clone="mh2\_4614"  
 /clone\_lib="msh"  
 /note="Environmental isolate. Whole genomic shotgun library."

**BASE COUNT** 240 a 394 c 318 g 371 t

**Alignment Scores:**  
 Pred. No.: 9.61e-84 Length: 1323  
 Score: 820.50 Matches: 171  
 Percent Similarity: 81.67% Conservative: 34  
 Best Local Similarity: 68.13% Mismatches: 37  
 Query Match: 43.97% Indels: 12  
 DB: 29 Gaps: 2

US-09-921-992-78 (1-372) x B575759 (1-1323)

115 Gtugtargtgaatgmetvalaaspccysalaaargaspysasnleproileargtll 134  
 :::  
 86 CAGCCCGGCGAAGCGCGGTGCGACCGCGCGCGCGCAACATCCGATCGGTATC 145  
 135 Gtlyaaanaaiaaglyserleuglnulysapleugngulystyrglygluprothpro 154  
 :::  
 146 GCGGCAATGCCGCTGCTCGTAAGAAAGACCTGCAAGAAATACGGGAAACCGACCCG 205  
 :::  
 155 Gtalaaleuagluaserlaametarghisvalasphisleuaspargleuasphasp 174  
 :::  
 206 GAAGCCCTCGCATGCGCATGCGCATGCGCATGCGCATGCGCATGCGCATGCGCATG 265  
 175 Glnphelyvalservallysalaaseraspyalpheleualavalglusertyargleu 194  
 :::  
 266 AACTTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 325  
 195 Leualalysglniliaspplnproleuhsleuglyilethrgualaglglvalaarg 214  
 :::  
 326 CTGGCGAAGCAATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 385  
 215 SerglalavalysaserlaaleglleuglyleuSerleuSerleuSerleuSerleu 234  
 :::  
 386 TCGGCAAGGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 445  
 235 Thrleualgvalserleualalalasprowalglugluilelyvalaglphaspile 254  
 :::  
 446 ACCATCCGATTCCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 505  
 255 Leulyserleuargileargserargyleanphelealealacysprothrcysser 274  
 :::  
 506 CTCAGTCCCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 565  
 275 Argnglglupheaspvalilegllythryalalalalalalalalalalalalalal 294  
 :::  
 566 CGGCAAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 625  
 295 lletnpprometaspvalserlleilegllycysvalavalasnlyprogllylualaleu 314

Db 626 CTGTGCGCATGAGCAGCGCGCGCTAATCGT-TGCGTGTAAACGCTCGGCGAAGCCAA 684  
 :::  
 Oy 315 valserthleuglyvalthrglygylasulysserleuglyleuargylval 334  
 :::  
 Db 685 GAGCCCATGCTGCTC-ACCGGCGGCC-TCGCAACTGCTGTATTCGACGGGAAAG 742  
 335 Arg---lysapargleuaspaasnasuametilaspplnleuglnualaargtarg 353  
 743 CCGGTGCACAAATCTGGCCCAACAGCAACCTGTGTGACCAAGCTT----- 787  
 Oy 354 Alalysalaserleuasaspglualaargarg 364  
 :::  
 Db 788 -----GTACGTTGTATTCGCCCAAAAG 811

**RESULT 2**  
 CB669733 828 bp mRNA linear EST 09-APR-2003  
 LOCUS OSJNE02H05.1 OSJNE Oryza sativa (japonica cultivar-group) cDNA  
 DEFINITION  
 OSJNE02H05.1 OSJNE Oryza sativa (japonica cultivar-group) cDNA  
 clone OSJNE02H05.5', mRNA sequence.  
 CB669733  
 CB669733.1 GI:29673458  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### COMMENT

Contact: Rod Wang  
 Arizona Genomics Institute  
 University of Arizona  
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
 85721-0088, USA  
 Tel: 520 626 3967  
 Fax: 520 621 9288  
 Email: http://genome.arizona.edu  
 PCR primers  
 FORWARD: gta aaa cga cgg cca gtc  
 BACKWARD: gga aac agc tat gac cat g  
 Plate: 02 row: H column: 05  
 Seq primer: gta aaa cga cgg cca gtc.  
 Location/Qualifiers

**FEATURES**  
 source  
 1..828  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultur="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="OSJNE02H05"  
 /tissue\_type="Leaf"  
 /dev\_stage="3 week"  
 /lab\_host="DH10B"  
 /clone\_lib="OSJNE"  
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2: XhoI; 24 hrs after inoculation with Rice Blast (70-15)";

**BASE COUNT** 231 a 155 c 217 g 224 t 1 others

#### ORIGIN

**Alignment Scores:**  
 Pred. No.: 3.95e-33 Length: 828  
 Score: 379.00 Matches: 91  
 Percent Similarity: 55.27% Conservative: 40  
 Best Local Similarity: 38.40% Mismatches: 72  
 Query Match: 20.31% Indels: 34  
 DB: 14 Gaps: 6

US-09-921-992-78 (1-372) x CB669733 (1-828)

Fax: 520 621 9288	
Email: <a href="http://genome.arizona.edu">http://genome.arizona.edu</a>	
PCR Primers	
FORWARD: gta aac cga cgg cca gtc	
BACKWARD: gga aac agc tat ggc cat g	
Plate: 03 Row: 0 Column: 03	
Seq primer: gta aac cga cgg cca gtc.	
Location/Qualifiers	
FEATURES	
source	1. 853
	/organism="Oryza sativa (indica cultivar-group)"
	/mol_type="mRNA"
	/cultivar="IR36"
	/db_xref="taxon:39946"
	/clone="OSIRPa03003"
	/tissue_type="leaf"
	/dev_stage="3 week"
	/lab_host="DH10B"
	/clone_lib="OSIRPa"
	/note="Vector: pBluescript II KS +, site_1: EcoRI, site_2:
	XhoI; lesion mimic SPL 11"
BASE COUNT	222 a 168 c 223 g 240 t
ORIGIN	
Alignment Scores:	
Pred. NO.:	1,35e-29 Length: 853
Score:	348.50 Matches: 81
Percent Similarity:	56.93% Conservative: 34
Best Local Similarity:	40.10% Mismatches: 58
Query Match:	18.68% Indels: 29
DB:	14 Gaps: 5
US-09-921-992-78 (1-372) x CB619369 (1-853)	
OY	74 LeuilelsglnclnlnvalnsnvalProleuvalAlaasplleHisPhasprtyrArglle 93
Db	8 CTTCGTCCAGAAATTACACACATCCCCCTAGCGGTGATATTCATTTGCCCGACAGTT 67
OY	94 AlaLeuysvalAlaGlnurgrglYvalAaPcysleuAcrglleAsnProglYAsnlllely 111
Db	68 GCTTAAAGAGTGGCTGAATGC---TTTGACAAAAATTCGTGTACACCCAGGAAATTTGCT 122
OY	114 AsnGlu----- 111
Db	125 GATCGCCGTGCCCAATTTAGCAGAGCTTGAAATACTGAAGATGATTCAAAAAGAGCTT 184
OY	116 GluArglleArgmetval-----ValaPcysAlaArgAspLyAsnlllePro 133
Db	185 GACCAATTCGACAGAGGTCTTCTCCCGGTGGTGGAATAAGCAGACAGTATGAAGAGCA 244
OY	132 lleArglleglYvalAsnAlaGlyserleuGlnulysAspLeuGlnulysTyrGlyGlu 151
Db	245 ATCCGTATATAGGAACAATATGATGGAAGTGTGTGTGACCCGCAATAAAGATTTACTATGGTAT 304
OY	152 ProthProGlnAlaLeuLeuGlnuserAlaMetArgHisValAspHisLeuAspArgleu 177
Db	305 ---TCGCCAGCGGAATGGTTGAGTCTGCTTGGAAATTTGCCAGAGATCTTCCGGAACTG 361
OY	172 AsnPhaspglnPheLysvalSerValysAlaSerAspValPheLeuAlaValGlnSer 191
Db	362 GACTTCATTAACATCTTGTCTTTTCATTAAGAAAGCAAGTAAACCTGTATCATGTGTCCAAACA 421
OY	192 TyrArgLeuLeuAlaLysGlnlle-----AspGlnProLeuHisLeuGly 206
Db	422 TATCGCTTCCTGTGTAGCAGAAATGTGATTAACCTAGAGGTGGATGTATCTTGCACCTTGGGA 481
OY	207 llethGlnAlaGlyGlyAlaArgserGlyAlaValLysSerAlaAlaGlyLeuGlyLeu 226
Db	482 GTTACGAAGAGCTGGAGAGGCTGAAGAATGGAGATGAAGTGTGCATTTGGCATTTGGACACA 541
OY	227 LeuLeuSerGlnGlylleGlyAspPhrLeuArgValSerLeuAlaAlaAspProValGlu 246
Db	542 CTTCGTATGATGAGCTTGGCGATTCATTCGCTGCTCCCTACGAGCAACCTGAGGAA 601



OY	247	Glut1e 248		
Db	602	GAGATT 607		
RESULT 4				
LOCUS	AY104363			
DEFINITION	zea mays PCOL16062 mRNA sequence.	1442 bp	mRNA	Linear
ACCESSION	AY104363			
VERSION	AY104363.1	GI:21207441		
KEYWORDS				
SOURCE	zea mays			
ORGANISM	zea mays			
REFERENCE				
AUTHORS	Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitstilt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.			
TITLE	Maize Mapping Project/Dupont Consensus Sequences for Design of Overgo Probes			
JOURNAL	Unpublished (2002)			
REFERENCE	2 (bases 1 to 1442)			
AUTHORS	Coe,E.H.			
TITLE	Direct Submission			
JOURNAL	Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA			
COMMENT	If you are interested in getting corresponding physical clones, these are publicly available from zmdb and may be found by BLAST searching at MSU, maizegap.org; Zmdb, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize CDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from Zmdb: www.zmdb.iastate.edu.			
FEATURES	Location/Qualifiers			
Source	1..1442			
	/organism="Zea mays"			
	/mol_type="mRNA"			
	/db_xref="MaizeDB:636753"			
	/db_xref="taxon:4577"			
	/clone_lib="Maize Mapping Project/Dupont Consensus Library"			
	/note="this sequence is part of a project of EST assemblies resulting from the application of public configs to seed Dupont configs; this resource was assembled by Dupont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"			
BASE COUNT	381 a	282 c	374 g	399 t
ORIGIN	6 others			
Alignment Scores:				
Pred. No.:	9,68e-27	Length:	1442	
Score:	326.50	Matches:	99	
Percent Similarity:	49.66%	Conservative:	46	
Best Local Similarity:	33.90%	Mismatches:	77	
Query Match:	17.50%	Indels:	71	
DB:	11	Gaps:	11	
US-09-921-992-78 (1-372) x	AY104363 (1-1442)			
OY	65	Thrmetsapaalaaaglaalaphelys-----	LeuilelysglgnlValasn	80
Db	3	ACGAGGAGGAGCGATGCTCTTGATGATCAAGAACACTCGTGCACAGAAATTCACAC	62	
OY	81	ValProleuValAlaAspIleHisshespyrAryllleAlaLeuLysValAlaGluYr	100	
Db	63	ATTCACATGAGGCGCGATATTCATTTCTCCTCAGGATGAGCCCTAAGGGTGGCAGATGT	122	
OY	101	GIyValaAspCyLseAArgIleAsnProGlyAsnIleGlyAsnGlu-----	115	
Db	123	---TTTGACAAATTCGTGTGATGACCCAGAGAAATTTTGCTGATCGTCGTCATATTTGGA	179	

Oy		116	-----ValapscysalaArgAspLysasnIleProileArgIleGlyValAsnAla	121
Db		180	AAGCTGAATATACTACGACGAGCTACCAGAAAGAGCTAAGCGATTATGGAAAGGTGTTT	239
Oy		122	-----ValapscysalaArgAspLysasnIleProileArgIleGlyValAsnAla	138
Db		240	TTCCTCATTAAGTAGGAATGACAGTATGGAAGACATGCAGATAGCAACAATCAT	299
Oy		139	GlySerLeuGluAsnAspLeuGlnGluIuLysTygGlyLubProHrProGlnAlaLeuLeu	158
Db		300	GGTAGCTTTCTTGACCCCATPATGACGTACTACTGATAT--TCTCCAGGGCAATGGTT	356
Oy		159	GluSerAlaMetArgHisValAspHisLeuAspArgLeuAsnPhaSpGlnPheLysVal	178
Db		357	GAGCTGCTTGGAAATTTGCTAGAATATGTGGAAAGTTGGACTCCTCCAACTTTGATT	416
Oy		179	SerValLysSalaseraSpsValPheLeuAlaValGluSerTyrrArgLeuLeuAlaLysGln	198
Db		417	TCGATGAAGCTAGTAACCCCTGTATTAGTGCACACATATGCCCTGTGTGGCAGAA	476
Oy		199	Ile-----AspGlnProLeuHisLeuGlnLylIethrGluAlaGlyVala	213
Db		477	ATGATAAACCTAGSATGGAGTATTATCCCTGTGACTGTGGGTGTACAGAACTGGAGAGGT	536
Oy		214	ArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuLeuSerGlnGlyIleGly	233
Db		537	GAAATATGA-AGGATGAATCTGCTATTGGCATTTGGGACACTGCTAATGATGATTGGGT	595
Oy		234	AspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle-----	248
Db		596	GATACAAATCCGTCTCTCCCTCACAGAACACACAGAAAGAGATTGATCCTGGCCAAAG	655
Oy		249	-----LysValGlyPheAspIleLeuLysE	257
Db		656	TTGGCAAATCTTGGGAGCGACGCGCAAACTTCGAAATTTGGGGGGCCCCATTGAAAG-	713
Oy		257	rLeuArGlrArgSerArGlyIleAsnPheIleAlaCySProThrcysSerArgGlnI	277
Db		714	-----AAAAGCACAGCGCTTATTTTATTTCCAGC-----	743
Oy		277	uPheaSpsValIleGlyThrValAsnAlaLeu--GluGlnArgLeuGluAspIleIeth	296
Db		744	-----GTAGAGAGTGCATATGCTTTTGACGAGAAGGAGGGTGAAGAAATTGACTAC	793
Oy		296	rPrometAspValserIleIleGlyCysValVal	307
Db		794	AGAAATG-----TCTGTGATCGTG	812
<hr/>				
RESULT 5	BZ569685	1268 bp	DNA linear	SSS 17-DEC-2002
LOCUS	pacs2-164_970.s1 pacs2-164 Pseudomonas aeruginosa genomic clone			
DEFINITION	pacs2-164_970, genomic survey sequence.			
ACCESSION	BZ569685			
VERSION	BZ569685.1	GI:27204646		
KEYWORDS	SSS.			
SOURCE	Pseudomonas aeruginosa			
ORGANISM	Pseudomonas aeruginosa Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas. 1 (bases 1 to 1268) Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V. Whole-genome-sequence variation among multiple isolates of Pseudomonas aeruginosa library J. Bacteriol., (2002) In press Contact: Chris K. Raymond Genome Center University of Washington Box 352145, Seattle, WA 98105-2145, USA Tel: 2062216954 Fax: 2066857244 Email: craymond@den.washington.edu			





```

Db      71 TGC---TTTGACAAATACGTCTCAATCCTGAAACCTTGCTGACAGCGAGCCAGTTT 127
QY      115 -----
Db      128 GAGCAATTAGAGTACAGGAAAGATGACTATCAGAAAGACTCGACCAATTATGAGAGCTT 187
QY      118 TleatmetValValaspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsn 137
Db      188 TTTACACCATTTGTCGAAATAATGTAGAAGATGACGTGCAATCGCATTCGGACAAAC 247
QY      138 AlaGlySerLeuGluLysAspLeuGlnIleuLysTyrGlyLysProIleProGlnAlaLeu 157
Db      248 CATGGAGACCCCTTCAGATCGCATTTAGAGCTATTATGAGGAC--TCGCTTAGGGGAATG 304
QY      158 LeuGlnSerAlaMetArgHisValAspHisLeuAspArgLeuAsnPhaSpGlnPheLys 177
Db      305 GTAGAAATACGATTTGAGTTTGCAAGAAATTTGTAGAAAGTTGACTTTCCACAAATTCGTC 364
QY      178 ValSerValLysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuAlaLys 197
Db      365 TTCTCAATGAAGAAGCTAGCAATCCATGACTATGTTTACGCGGATCGCTTCTTACTAGCT 424
QY      198 GlnIle-----AspGlnProLeuHisLeuGlyIleIleThiGlnAlaGly 212
Db      425 GAGATGATGTTCACAGGAGTGGGACTATCCCTTACCTTGAGGAGTTCAGAACCTGGTGA 484
QY      213 AlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuSerGlnIle 232
Db      485 GGTGAGATGAGAGAGAGTGAAGTCTGATGATGCTATGTCACACTTCTTCAGAGATGTTG 544
QY      233 GlyAspThrLeuArgValSerLeu 240
Db      545 GGTGATACAAATACGGGTTCCCTA 568

RESULT 7
LOCUS   BH235010              720 bp      DNA      linear      GSS 01-JAN-2002
DEFINITION PH.05.x Spiroplasma kunkelii H Spiroplasma kunkelii genomic clone
ACCESSION BH235010
VERSION   BH235010.1 GI:18030478
KEYWORDS  GSS.
SOURCE    Spiroplasma kunkelii
ORGANISM  Spiroplasma kunkelii
REFERENCE 1 (bases 1 to 720)
AUTHORS   Hogenhout,S.A.
TITLE      Genomic sequences from Spiroplasma kunkelii strain M2
JOURNAL   Unpublished
COMMENT    Contact: Hogenhout SA
           Department of Entomology
           The Ohio State University-OARDC
           120 Thorne Hall, 1680 Madison Avenue, Wooster, OH 44691, USA
           Tel.: 330 263 3730
           Fax: 330 263 3686
           Email: hogenhout.1@osu.edu
           Class: HindIII fragments.
FEATURES
  source             1..720
                     /organism="Spiroplasma kunkelii"
                     /mol_type="genomic DNA"
                     /strain="M2"
                     /db_xref="taxon:47834"
                     /clone="PH.05.x"
BASE COUNT      255 a          96 c          132 g          233 t          4 others
ORIGIN
BASE COUNT      255 a          96 c          132 g          233 t          4 others
ORIGIN
Alignment Scores: 1.32e-25      Length: 720
Prid. No.: 313.00      Matches: 67
Score:

```

```

Percent Similarity: 62.20%      Conservative: 35
Best Local Similarity: 40.85%      Mismatches: 61
Query Match: 16.77%      Indels: 2
DB: 28      Gaps: 0

US-09-921-992-78 (1-372) x BH235010 (1-720)
QY      190 GluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThiGln 209
Db      31 CAAGCTTATACCTTAGCTAGTAAGTAAGTAAGTAATTCCTTCATCTAGGAGTGTCT-GAA 89
QY      210 AlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuSer 229
Db      90 GCTGTAAGTATCATCATACCGAACAANNAATCATGTAAGTGTGTGTCAACCGCTTCTTTT 149
QY      230 GlnGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluIleLys 249
Db      150 AATGTAATTTGGATACCATTCGAAATTTAGTTATATCACTGATCCATACGACAGGTGGA 209
QY      250 ValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIleAla 269
Db      210 GTTGGTAAACGAAATGTTAAATTCATAGCACTTTATGATTAACATTTGTGATTAATTCCT 269
QY      270 CysProThrCysSerArgGlnGluIlePheAspValIleGlyThrValAsnAlaLeuGln 289
Db      270 TCCCAACATGTCGTGCTTGAATATGACCTTTCCCGCTGTGTAAGAAATTAAGACAA 329
QY      290 ArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsnGly 309
Db      330 TATACAAAGAATTTAAATTAATTCATTAATAAATGCAATTTAGTGTGTGTTAATGCT 389
QY      310 -ProGlyGluAlaLeuValSerThrLeuGlyAlaThrGlyLysAsnLysSerGlyLe 329
Db      390 CCCAGCAAGAGCTAAACAACGATGATGAGGATTCCTGCTGTGTAATAAATGTGTGAATAT 449
QY      329 uTyrGluAspGlyValAlaArgLysAspArgLeuAspAsnAspMetIleAspGlnLeu 349
Db      450 TTTTAAAGTAAATTAATTTAATTAATCATGTAACAAAGAAATCTTTGTCAGAAATTA 509
QY      349 uAlaArgIle 352
Db      510 ACTATTAAAT 519

RESULT 8
LOCUS   AZ570993/C              658 bp      DNA      linear      GSS 15-MAY-2001
DEFINITION 281PVD07 PV MBN #30 Plasmodium vivax genomic 3', genomic survey
ACCESSION AZ570993
VERSION   AZ570993.1 GI:13982637
KEYWORDS  GSS.
SOURCE    Plasmodium vivax (malaria parasite P. vivax)
ORGANISM  Plasmodium vivax
REFERENCE 1 (bases 1 to 658)
AUTHORS   Carlton,J.M.,R. and Dame,J.B.
TITLE      The Plasmodium vivax and P. berghei gene sequence tag projects
JOURNAL   Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)
COMMENT    Contact: Dame JB
           Dept. of Pathobiology, College of Veterinary Medicine
           University of Florida
           2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
           Tel: 352 392 4700
           Fax: 352 392 9704
           Email: damej@mail.vetmed.ufl.edu
           Seq primer: M13(-20) forward
           Class: shotgun.
FEATURES
  source             1..658
                     /organism="Plasmodium vivax"
                     /mol_type="genomic DNA"
                     /strain="Salvador I (Collins, W. 1972. J. Parasitol. 69,
                     497-598)"

```

```

/gb_xref="taxon:5855"
/dev_stage="asexual blood forms"
/lab_host="Salimri holivensis"
/clone.lib="pv MBN #30"
/notes="Vector: pBluescript SK(+) vector DNA, phagemid
excised from lambda ZAP; Site 1: EcoR V; Site 2: EcoR V;
Host leukocytes were extracted from P. vivax infected
blood using the following methods: first, infected blood
was activated by the addition of 0.5 ml of ADP (40mg/ml)
per 10 ml blood. Then blood was passed over a column of
filter washed 0.1 mm glass beads, then through a plasmidpur
filter, followed by passage through a column of pre-wet
Whitman Cfil powder (1:2 ratio volume of blood to Cfil),
and finally centrifuged through a 50% Percoll density
cushion. Purified DNA was digested with mung bean nuclease
in the presence of 44% formamide at 50°C as described
(Venick, R.D., Imberski, R.B., and McCutchan, T.F. 1988.
Nucleic Acids Research 16:6883-6896). Digested DNA was
blunt-ended using T4 DNA Polymerase and size fractionated
over a Sepharose CL-2B column. Fractions in the size range
500bp-4kb were ligated into the Eco RV site of pBluescript
SK(+), and E. coli XL-10 Gold transformed with the
ligation mixture."
BASE COUNT      178 a      150 c      119 g      206 t      5 others
ORIGIN

```

```

Alignment Scores:
Pred. No.:      5,01e-25      Length:      658
Score:          307.50      Matches:      79
Percent Similarity: 53.85%      Conservative: 33
Best Local Similarity: 37.98%      Mismatches: 61
Query Match:    16.48%      Indels:      35
DB:              Gaps:      6

```

US-09-921-992-78 (1-372) x AZ570993 (1-658)

```

Oy 73 LysleuileLysGlnGlnValValProleuValAlaAspIleHisPheAspTyrArg 92
Db 642 AAATGGAAGCTCAAAATATACATTCCTGTTAGACGACATCAATTTAAATCCAA 583
Oy 93 IleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsnProGlyAsnIle 112
Db 582 ATTCTCTCATGCAAGAGATGTA---TTTGATTAATAATACGATTAACCCAGGAAATTAC 526
Oy 113 GlyAsnGluGluArg----- 117
Db 525 GTCATGTGGAGNAAGAAATGATCAATTAACAACGAGAGAGANTTCATCAA 466
Oy 118 -----IleArgMetValValAspCysAlaArgAspLys 128
Db 465 GCGAAATTAATTAATTCAGNAGAGTTCATCCCTGATTAAGAGAGG---TTA 409
Oy 129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnLys 148
Db 408 AATAGACCATGACCAATCGCACAAATACGCGCTCTATCTTAGGACTCTCTTTC 349
Oy 149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
Db 348 TATGGGAGAT---ACCCNTTAGATGATGCGGCTTTAAATTTCCGATTTGTGT 292
Oy 169 AspArgLeuAsnAspHisAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
Db 291 GTACAGAAATTAACCTTCTCAATGTGTTTCTCCATGAAGGCTTCAACCGGATATATAG 232
Oy 189 ValGluSerTyrArgLeuLeu---AlaLysGlnIleAspGln----- 201
Db 231 ATACAGTCGATAGGCTACTCGTGGCGGACCAATATGAAGATGGGCAACGGGTACTA 172
Oy 202 ---ProLeuHisLeuGlyIleThrGluAlaGlyLysValArgSerGlyAlaValLysSer 220
Db 171 TTCCCACTGCAATTTGGCGCTTACCGAAGAGGTTTGGGATTAACGAGATTAATATCC 112
Oy 221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240

```

```

Db 111 TATTGGGAGATAGTTCCTGCTGACAGCGGATAGAGACACATCAGATTTCTTA 52
Oy 241 AlaAlaAspProValGluGluIle 248
Db 51 ACTGAGGACCTTGGGAGGAAATTA 28

```

```

RESULT 9
LOCUS
DEFINITION
CB628479 852 bp mRNA linear EST 08-APR-2003
OS:Oryza sativa (indica cultivar-group)
Gene: Oryza sativa (indica cultivar-group)
clone Oryza sativa (indica cultivar-group)
CDNA
CB628479.1 GI:29623468
VERSION
KEYWORDS
SOURCE
ORGANISM
Oryza sativa (indica cultivar-group)
Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 852)
Jantanasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aaa cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 04 row: B column: 06
Seq primer: gta aaa cga cgg cca gtc.
Location/Qualifiers
1. 852
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="IR36"
/gb_xref="taxon:39946"
/clone="OSIIE04B06"
/tissue_type="leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone.lib="OSIIE04B06"
/notes="Vector: pBluescript II KS+, Site 1: EcoRI, Site 2:
XhoI; 24 hrs after inoculation with Rice Blast (P06-6-3)"

```

```

BASE COUNT      233 a      172 c      223 g      224 t
ORIGIN

```

```

Alignment Scores:
Pred. No.:      2.69e-21      Length:      852
Score:          276.50      Matches:      72
Percent Similarity: 53.43%      Conservative: 37
Best Local Similarity: 35.29%      Mismatches: 66
Query Match:    14.82%      Indels:      30
DB:              Gaps:      5

```

US-09-921-992-78 (1-372) x CB628479 (1-852)

```

Oy 8 GlnaArgGlySerThrArgIleTyrAlaGlnValProIleGlyAspGlyAlaPro 27
Db 238 AGGAGGAGAAAAAAGCCGACGTGATGTTGGGATGCGCAGTGGAGATGATCAATCC 297
Oy 28 IleAlaValGlnSerMetThrAsnThrArgThrAspValGluAlaThrValAsnGln 47
Db 298 ATTAGGATTCAGACTATACCACTCGATACCAAGAGATGTTGCTAAACCGCTGAGAG 357

```





```

QY 90 AsptYrArgIleAlaLeuValAlaGluTyrGlyValAspCysLeuArgIleAsnPro 109
Db 13 GACAGGGGAGCCAGGTTGAGCAATAGATACACAGAGATACATATCCAGAAAGAACTC 72
QY 110 GtAsnIleGlyAsnGluGluArgIleArgMetValAlaSpCysAlaArgAspLysAsn 129
Db 73 GAGCAATATTTTGGAGGAGTTTACACCATTTGGTGGAAAAATGATAGAGATATGGA 126
QY 130 IleProIleArgIleGlyValAlaAsnAlaGlySerLeuGluLysAspLeuGluLysTyr 149
Db 127 CGTGCATATGCGCATTTGGGACAAACCATGGAGCCCTTCAGATGCAATATGAGCTATTA 186
QY 150 GlyGluProThrProGlnAlaLeuGluSerAlaMetArgHisValAspHisLeuAsp 169
Db 187 GGGGAC---TCGCTTGGGGAATGATAGATACAGATTGAGTTGACAGAAATTTGTAGA 243
QY 170 ArgLeuAsnPhesPglNphelysValSerValLysAlaSerAspValPheLeuAlaVal 189
Db 244 AAGTTGACTTTCAGATTTTCGCTCTCTCAATGAAAGCTAGCAATCAGTATGATGCTT 303
QY 190 GluSerTyrArgLeuAlaLeuAlaLysGlnIle-----AspGlnProLeuHis 204
Db 304 CAGGCGTATGCGCTTGGTGTAGCTGATGATGTTACAGAGATGGAGACTATCTTTACAC 363
QY 205 LeuGlyIleThrGlnAlaGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeu 224
Db 364 TTGGGAGTTACTGACGCTGTGTGAAGGTGAGATGGAAGATGAACTGCAATCGGTAT 423
QY 225 GlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaIleAspPro 244
Db 424 GGCACACTCTTCAGAGATGTTGGTGTATACATATCGGGTTTCCCTAACCTAGACTCA 483
QY 245 ValGluGluIle 248
Db 484 GAAGAGAGATA 495

RESULT 13
AU186794 standard; RNNY EST; 489 BP.
XX AU186794;
AC AU186794;
SV AU186794.1
DT 23-MAY-2003 (Rel. 75, Created)
DT 23-MAY-2003 (Rel. 75, Last updated, Version 1)
XX Porphyra yezoensis cDNA, Clone: PF004D08_r, 5' end.
DE Porphyra yezoensis cDNA, Clone: PF004D08_r, 5' end.
XX EST (expressed sequence tag).
KW Porphyra yezoensis
OS Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
XX [1]
RN 1-489
RA Asamizu E., Nakamura Y.;
RT Submitted (15-JUN-2001) to the EMBL/GenBank/DBJ databases.
RL Erika Asamizu, Kazusa DNA Research Institute, The First Laboratory for
PL Plant Gene Research, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
RL (E-mail: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/,
Tel: 81-438-52-3935 (ex. 2336), Fax: 81-438-52-3934)
XX [2]
RN Asamizu E., Nakajima M., Kitade Y., Saga N., Nakamura Y., Tabata S.;
RT "COMPARISON OF RNA EXPRESSION PROFILES BETWEEN THE TWO GENERATIONS OF
RT PORPHYRA YEZOENSIS (RHODOPHYTA), BASED ON EXPRESSED SEQUENCE TAG FREQUENCY
RT ANALYSIS";
RL J. Phycol. 0:0-0(2003).
XX Key Location/Qualifiers

```

```

FH source 1. 489
FT /db_xref="taxon:2788"
FT /mol_type="mRNA"
FT /organism="Porphyra yezoensis"
FT /clone="PF004D08_r"
FT /dev_stage="sporophytes"
FT /strain="Tu-1"
XX
SQ Sequence 489 BP; 100 A; 136 C; 170 G; 83 T; 0 other;
Alignment Scores:
Pred. No.: 3.42e-20 Length: 489
Score: 264.00 Matches: 68
Percent Similarity: 57.148 Conservative: 24
Best Local Similarity: 42.248 Mismatches: 48
Query Match: 14.158 Indels: 22
DB: Gaps: 5
US-09-921-992-78 (1-372) x AU186794 (1-489)
QY 104 CysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValAlaSp 123
Db 14 TGTTTGAGA---AGCCCAAGAGAGGCGCGACGAGTACTCGGACGAGGATTCACACCGGC 70
QY 124 -----CysAlaArgAspLysAsn----- 129
Db 71 TCAAAGACAAGCTCAAGCGCCACCCCTCACACCACTGCTGTGTGCTCAAAAGACAGGCA 130
QY 130 IleProIleArgIleGlyValAlaAsnAlaGlySerLeuGluLysAspLeuGluLysTyr 149
Db 131 AGCCCATG-CGTATTTGGGTCACACCGGTCCTCGGCGGAGCCGATGATGTCACCTAC 189
QY 150 GlyGluProThrProGlnAlaLeuGluSerAlaMetArgHisValAspHisLeuAsp 169
Db 190 GGGGAC---ACCCCGCGGGGATGTCAGATCGGCGATGAGTGCATGATGTCGCCG 246
QY 170 ArgLeuAsnPhesPglNphelysValSerValLysAlaSerAspValPheLeuAlaVal 189
Db 247 GAGCTAGACTTTCACACACCTGCTGTTCATGTGAAGGCGTCCAAAGTCCAGTCAATGATC 306
QY 190 GluSerTyrArgLeuAlaLeuAlaLysGlnIle-----AspGlnProLeuHis 204
Db 307 GCAGCTTACCGCAGCTGCGCTGCGCTGACCGCTGAGGGGTACGATCACTCCATCCAC 366
QY 205 LeuGlyIleThrGlnAlaGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeu 224
Db 367 CTGGGTGTGACTGAGCGCGGCGAGCGGAGATGGGCGGATGGAAGCTCATGCTGATC 426
QY 225 GlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaIleAspPro 244
Db 427 GGCACGCTCTTGGGAGGAGCATGCGGACACGCTGCGGGTGACCTTGACAGAAACCCC 486
QY 245 Val 245
Db 487 ATG 489

RESULT 14
AM042702 LOCUS 601 bp mRNA linear EST 18-SEP-1999
DEFINITION ST33E04, Pine Triplex shoot tip library Pinus taeda cDNA clone
ACCESSION AM042702, mRNA sequence.
VERSION AM042702.1 GI:5903147
KEYWORDS EST.
SOURCE Pinus taeda (loblolly pine)
ORGANISM Pinus taeda
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE 1 (bases 1 to 601)
AUTHORS Whetten,R.W., Kinlaw,C.S., Retzel,E. and Sederoff,R.R.
TITLE The Pine Gene Discovery Project
JOURNAL Unpublished

```

## COMMENT

Contact: Ross Whetten  
Forest Biotechnology Group  
North Carolina State University  
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh  
NC, 27695-8008  
Tel: 919-515-7800  
Fax: 919-515-7801  
Email: rosswhetten@ncsu.edu  
Seq primer: 5' lambda Triplex2 Sequencing Primer.  
Location/Qualifiers

## FEATURES

1. 601

/organism="Pinus taeda"

/mol\_type="mRNA"

/db\_xref="taxon:3352"

/clone="ST3E04"

/lab\_host="E. coli BM25.8"

/note="Organ: shoot tips; Vector: Lambda Triplex; Site: 1: SfiI (A); Site: 2: SfiI (B); Shoot tips (approx. 2 cm from apex) were collected during the spring, frozen and used for mRNA isolation. The SMART-PCR method (Clontech) was used to prepare a library from 1 ug total RNA, using the Lambda Triplex vector. Plasmid subclones in pTriplex were recovered by cre-lox excision in E. coli strain BM25.8 and sequenced from the 5' end."

## BASE COUNT

162 a 103 c 166 g 165 t 5 others

## ORIGIN

## Alignment Scores:

Pred. No.: 7.76e-20 Length: 601  
Score: 262.00 Matches: 57  
Percent Similarity: 62.30% Conservative: 19  
Best Local Similarity: 46.72% Mismatches: 40  
Query Match: 14.04% Indels: 6  
Gaps: 2

US-09-921-992-78 (1-372) x AM042702 (1-601)

QY 132 ILEATGIIIEGLYVALASNAAGLYSERLEUGLULYSASPLEUNGILULYSTYGYGLU 151  
DB 24 ATGCGAATTCGACAAACCATGGAAGCTCTCCGATCGTACTAGATTATATGGTGA 83  
QY 152 ProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeuAspArgLeu 171  
DB 84 ---TCTCCAGGGGTATGTGATGATTCGATTTGATTCGACGATTTGCCGAAGTTG 140  
QY 172 AspPheAspGlnPheLeuValSerValLysAlaSerAspValPheLeuAlaValGluSer 191  
DB 141 GGTTCATATATTTGTTGTTTCAATGAAAGCGAGATCCCTGACTCATGTTTCAGCA 200  
QY 192 TyrArgLeuLeuAlaLysGlnIle-----AspGlnProLeuHisLeuGly 206  
DB 201 TACCGTTTACTTGTGGCGAGATGATGTCAGAGATGGATTTATCCATTCGATTGAGA 260  
QY 207 ILEthrgLUALAGLYGLYALaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeu 226  
DB 261 GTTACGAAAGCTGTGAAGGTGAAGTGAACGATGATGATTCGATTCGATTCGAA 320  
QY 227 LeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaLysProValGlu 246  
DB 321 CTTTTCAGAGATGTTGGTGTGATCTATTCGATTCGATTCCTTACAGAACCTCCAGAG 380  
QY 247 GluIle 248  
DB 381 GAGATC 386

## RESULT 15

## LOCUS

BT551332 603 bp mRNA linear EST 14-NOV-2002  
BT551332 K. Sato unpublished cDNA library, strain H602 adult,  
DERIVATION heading stage top three leaves Hordeum vulgare subsp. spontaneum  
cDNA clone bah59b20 5', mRNA sequence.

## ACCESSION

BT551332

## VERSION

BT551332.1 GI:24969783

## KEYWORDS

EST.

## SOURCE

Hordeum vulgare subsp. spontaneum

## ORGANISM

Hordeum vulgare subsp. spontaneum

## REFERENCE

Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae

## AUTHORS

1 (bases 1 to 603)  
Sato, K., Salscho, D. and Takeda, K.

## TITLE

Barley EST sequencing project in NIG and Okayama Univ

## JOURNAL

Unpublished

## COMMENT

Contact: Tadashi Shin-I

## FEATURES

Center for Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tsuhin@genes.nig.ac.jp.  
Location/Qualifiers

## source

1. 603

/organism="Hordeum vulgare subsp. spontaneum"

/mol\_type="mRNA"

/strain="H602"

/db\_xref="taxon:77009"

/clone="bah59b20"

/tissue\_type="top three leaves"

/dev\_stage="adult, heading stage"

/clone\_lib="K. Sato unpublished cDNA library, strain H602 adult, heading stage top three leaves"

## BASE COUNT

156 a 112 c 169 g 165 t 1 others

## ORIGIN

## Alignment Scores:

Pred. No.: 1.02e-19 Length: 603  
Score: 261.00 Matches: 56  
Percent Similarity: 63.93% Conservative: 22  
Best Local Similarity: 45.90% Mismatches: 38  
Query Match: 13.99% Indels: 6  
Gaps: 2

US-09-921-992-78 (1-372) x BT551332 (1-603)

QY 132 ILEATGIIIEGLYVALASNAAGLYSERLEUGLULYSASPLEUNGILULYSTYGYGLU 151  
DB 42 CTGCGTATCGGAACAAATCATGTGATGTTCTGTGACGGAATTAAGTACTATGATGAT 101  
QY 152 ProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeuAspArgLeu 171  
DB 102 ---TCTCCAGGGGTATGTGATGATTCGATTTGATTCGACGATTTGCCGAAGTTG 158  
QY 172 AspPheAspGlnPheLeuValSerValLysAlaSerAspValPheLeuAlaValGluSer 191  
DB 159 GACTTCATTAACCTTTGATTTTCAATGAAAGCAAGTACCTGTTGATGATTCGAA 218  
QY 192 TyrArgLeuLeuAlaLysGlnIle-----AspGlnProLeuHisLeuGly 206  
DB 219 TATGCCCTGCTTGAAGCGAATGATTAACCTTGATGATGATTCCTTGCATTCGGA 278  
QY 207 ILEthrgLUALAGLYGLYALaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeu 226  
DB 279 GTTACGAAAGCTGTGAAGGTGAAGTGAACGATGATGATTCGATTCGATTCGAA 338  
QY 227 LeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaLysProValGlu 246  
DB 339 CTTTTCAGAGATGTTGGTGTGATCTATTCGATTCGATTCCTTACAGAACCTCCAGAG 398  
QY 247 GluIle 248  
DB 399 GAAATC 404

Search completed: August 15, 2003, 01:35:09  
Job time : 2639 secs

Fri Aug 15 14:34:51 2003

us-09-921-992-78.rst

Page 12



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2003, 09:54:00 ; Search time 98 seconds  
(without alignments)  
\$79.547 Million cell updates/sec

Title: US-09-921-992-78

Perfect score: 1866

Sequence: 1 MHNQADIQRRKSTRIVYGVN.....RAKASQLEARRIDVQVEK 372

Scoring table: BLOSUM62

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP\_ARCHAEA:\*  
2: SP\_BACTERIA:\*  
3: SP\_FUNGI:\*  
4: SP\_HUMAN:\*  
5: SP\_INVERTEBRATE:\*  
6: SP\_MAMMAL:\*  
7: SP\_MHC:\*  
8: SP\_ORGANELLE:\*  
9: SP\_PHAGE:\*  
10: SP\_PLANT:\*  
11: SP RODENT:\*  
12: SP\_VIRUS:\*  
13: SP\_VERTEBRATE:\*  
14: SP\_UNCLASSIFIED:\*  
15: SP\_VIRUS:\*  
16: SP\_BACTERIAP:\*  
17: SP\_ARCHAEP:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1631.5	87.4	372	16 Q8DEZ8	Q8DEZ8 vibrio vuln
2	1584.5	84.9	371	16 Q8DEZ8	Q8DEZ8 shewanella
3	1226.5	65.7	366	16 Q8D1Y3	Q8D1Y3 wiggleswort
4	814.5	40.4	403	16 Q8G7Y6	Q8G7Y6 bifidobacte
5	753	39.6	362	16 Q8F82	Q8F82 coriobacte
6	739.5	32.2	402	16 Q8EUI6	Q8EUI6 mycoplasma
7	601.5	32.2	402	16 Q8DK70	Q8DK70 synecococc
8	525.5	28.2	460	16 Q8FJH5	Q8FJH5 leptospira
9	508	27.2	420	16 Q8FTT2	Q8FTT2 bruceella su
10	471	25.2	740	16 Q8GZK6	Q8GZK6 lycopersico
11	463	24.8	741	16 Q8RXG8	Q8RXG8 arabidopsi
12	462	24.8	716	10 Q9F59	Q9F59 arabidopsi
13	462	24.8	740	10 Q8GZK7	Q8GZK7 arabidopsi
14	457	24.5	741	10 Q8LP04	Q8LP04 arabidopsi
15	455.5	24.4	746	16 Q8KG23	Q8KG23 chlorobium
16	446	23.9	824	5 Q9BJX5	Q9BJX5 plasmodium

17	446	23.9	824	5 Q8ICU7	Q8ICU7 plasmodium
18	136	7.3	803	16 Q8R927	Q8R927 thermomane
19	122.5	6.6	507	17 Q8R2P6	Q8R2P6 pyrococcus
20	119	6.4	535	16 Q8D1J2	Q8D1J2 synecococc
21	114.5	6.1	588	22 Q8V5Q9	Q8V5Q9 actinobact
22	114	6.1	524	17 Q27773	Q27773 methanobac
23	112.5	6.0	686	17 Q9UXG1	Q9UXG1 sulfolobus
24	112.5	6.0	723	3 Q9P4D5	Q9P4D5 actinobact
25	111.5	6.0	344	16 Q8EQA9	Q8EQA9 oceanobact
26	111.5	6.0	588	2 Q68221	Q68221 anaplasm
27	111.5	6.0	588	2 Q8VSR0	Q8VSR0 actinobact
28	111	5.9	575	16 Q8VXK5	Q8VXK5 raistonia s
29	109	5.8	764	2 Q9RNG5	Q9RNG5 xanthomonas
30	108.5	5.8	449	16 Q8BUZ9	Q8BUZ9 xanthomonas
31	108.5	5.8	1000	17 Q8TPA1	Q8TPA1 methanocarc
32	106.5	5.7	284	16 Q9CHV6	Q9CHV6 lactococcus
33	106.5	5.7	798	2 Q9KHS7	Q9KHS7 pseudomonas
34	106.5	5.7	798	2 Q8KUR5	Q8KUR5 pseudomonas
35	106	5.7	748	17 Q57616	Q57616 methanococ
36	105.5	5.7	578	2 Q93NY9	Q93NY9 anaplasm
37	104.5	5.6	638	5 Q9BMO6	Q9BMO6 opisthorchi
38	104	5.6	606	16 Q8Y7W3	Q8Y7W3 listeria mo
39	104	5.6	1227	16 Q97K41	Q97K41 clostridium
40	103.5	5.5	537	16 Q8FL75	Q8FL75 escherichia
41	102.5	5.5	417	3 Q9C473	Q9C473 streptococ
42	102	5.5	546	16 Q8DZ63	Q8DZ63 streptococ
43	101.5	5.4	546	16 Q8DX35	Q8DX35 heliobact
44	101.5	5.4	564	5 Q44422	Q44422 heliobact
45	101.5	5.4	572	10 Q8GZK8	Q8GZK8 oryza sativ

## ALIGNMENTS

## RESULT 1

ID Q8DEZ8 PRELIMINARY; PRT; 372 AA.

AC Q8DEZ8; 01-MAR-2003 (TREMBLrel. 23, Created)

DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)

DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

GN VV10427.

OS Vibrio vulnificus.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;

OX NCBI\_TaxID=672;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CMCP6;

RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,

RT Choy H.E.;

RL "Complete genome sequence of Vibrio vulnificus CMCP6.";

DR EMBL; AE016798; AAC08950.1; -.

KW Complete proteome.

SQ SEQUENCE 372 AA; 40573 MW; 6C73C2B8819B0285 CRC64;

Query Match 87.4%; Score 1631.5; DB 16; Length 372;

Best Local Similarity 86.7%; Pred. No. 1.3e-112;

Matches 320; Conservative 29; Mismatches 19; Indels 1; Gaps 1;

QY 1 MHNQADIQRRKSTRIVYGVNPIGAGPIAVQSMNTFTTVEATVNOIKALERYGADIVR 60  
 DB 1 MHNESPIIRRRKSTRIVYGVNPIGAGPIAVQSMNTFTTVEATVNOIKALENGADIVR 60  
 QY 61 VSVPTMAAFAFKLIKQVNPPLVADIHEDYRIALKAERYGVDCRLRPNIGNIERIRM 120  
 DB 61 VSVPTMAAFAFKLIKQVNPPLVADIHEDYRIALKAERYGVDCRLRPNIGNIERIRS 120  
 QY 121 VVCCARDKNIPRIIGVAGSLEKDLQEKYGEPTPOLLESAMRHVDLNDQKRVSV 180  
 DB 121 VVCCARDKNIPRIIGVAGSLEKDLQEKYGEPTPOLLESAMRHVDLNDQKRVSV 180

QY 181 KASDVFLAVESYRLAKQIDOPHLGITTEAGGARSAGVKSALIGLLSGIGDTLAVSL 240  
 DB 181 KASDVFLAVESYRLAKQIDOPHLGITTEAGGARSAGVKSALIGLLSGIGDTLAVSL 240  
 QY 241 AADVVEIKVGFOLIKSRIRSGINFLIACPTCSROEDVIGTVNALQORLEDTITPMDV 300  
 DB 241 AADVVEIKVGFOLIKSRIRSGINFLIACPTCSROEDVIGTVNALQORLEDTITPMDV 300  
 QY 301 SIICGVANGPGEALVSTLGTGNGKSGLYEDGVR-KDRLDNNDMIDOLEARIRAKASQ 359  
 DB 301 SIICGVANGPGEALVSTLGTGNGKSGLYEDGVR-KDRLDNNDMIDOLEARIRAKASQ 359  
 QY 360 DEARRIDVQ 368  
 DB 361 DESNRIDIK 369

## RESULT 2

Q8EC32 PRELIMINARY: PRT: 371 AA.  
 AC Q8EC32;  
 DT 01-MAR-2003 (TREMBlrel, 23, Created)  
 DT 01-MAR-2003 (TREMBlrel, 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel, 23, Last annotation update)  
 DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.  
 GN ISP6 OR S03312.  
 OS Shewanella oneidensis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
 OC Alteromonadales; Shewanella.  
 RX NCBI\_TaxID=70863;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MR-1;  
 RC MEDLINE=22297686; PubMed=12368813;  
 RA Heidelberg J.F., Paulsen K.E., Nelson K.E., Gaidos E.J., Nelson W.C.,  
 RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,  
 RA Meyer T., Tsapin A., Scott J., Beaman M., Brinkac L., Daugherty S.,  
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
 RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,  
 RA Vadamthyan J., Weidman J.G., Imprial M., Lee K., Berry K., Lee C.,  
 RA Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A.,  
 RA Feldblyum T.V., Smith H.O., Venter J.C., Neale J.H., Fraser C.M.,  
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium  
 RT Shewanella oneidensis."  
 RT Nat. Biotechnol. 20:1118-1123(2002).  
 DR EMBL: AB015769; AAN56310.1;  
 DR TIGR: S03312;  
 KW Complete proteome.  
 SQ SEQUENCE 371 AA; 40603 MW; 024A994C6EC93906 CRC64;

Query Match 84.90% Score 1584.5; DB 16; Length 371;  
 Best Local Similarity 84.80% Pred. No. 3.8e-109;  
 Matches 313; Conservative 32; Mismatches 21; Indels 3; Gaps 2;

QY 1 MHNADIORRKRSTRIVGNVPIGDGAPIAVOSMTNRTTVDVATVNOIKALERVADIVR 60  
 DB 1 MHNADIORRKRSTRIVGNVPIGDGAPIAVOSMTNRTTVDVATVNOIKALERVADIVR 60  
 QY 61 MYNETIKRRPSTRIVGNVPIGDGAPIAVOSMTNRTTVDVATVNOIKALERVADIVR 60  
 DB 61 MYNETIKRRPSTRIVGNVPIGDGAPIAVOSMTNRTTVDVATVNOIKALERVADIVR 60  
 QY 61 VSVPTMDAAEAFKLIQSVNPLVADIHFDYRIALKAEEGVDCILRNIGNNEIRIM 120  
 DB 61 VSVPTMDAAEAFKLIQSVNPLVADIHFDYRIALKAEEGVDCILRNIGNNEIRIM 120  
 QY 121 VVDCARDKNIPRIGVNGSLKEDLOEKYGEPTPQALLSARHVDHLRNFDOFKVSY 180  
 DB 121 VVDCARDKNIPRIGVNGSLKEDLOEKYGEPTPQALLSARHVDHLRNFDOFKVSY 180  
 QY 181 KASDVFLAVESYRLAKQIDOPHLGITTEAGGARSAGVKSALIGLLSGIGDTLAVSL 240  
 DB 181 KASDVFLAVESYRLAKQIDOPHLGITTEAGGARSAGVKSALIGLLSGIGDTLAVSL 240  
 QY 241 AADVVEIKVGFOLIKSRIRSGINFLIACPTCSROEDVIGTVNALQORLEDTITPMDV 300  
 DB 241 AADVVEIKVGFOLIKSRIRSGINFLIACPTCSROEDVIGTVNALQORLEDTITPMDV 300

DB 241 AADVVEIKVGFOLIKSRIRSGINFLIACPTCSROEDVIGTVNALQORLEDTITPMDV 300  
 QY 301 SIICGVANGPGEALVSTLGTGNGKSGLYEDGVR-KDRLDNNDMIDOLEARIRAKASQ 359  
 DB 301 SIICGVANGPGEALVSTLGTGNGKSGLYEDGVR-KDRLDNNDMIDOLEARIRAKASQ 359  
 QY 360 DEARRIDVQ 368  
 DB 361 --ANRIGVK 367

## RESULT 3

Q8DIY3 PRELIMINARY: PRT: 366 AA.  
 AC Q8DIY3;  
 DT 01-MAR-2003 (TREMBlrel, 23, Created)  
 DT 01-MAR-2003 (TREMBlrel, 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel, 23, Last annotation update)  
 DE GCPE protein.  
 GN GCPE.  
 OS Wiglesworthia brevipalpis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Wiglesworthia.  
 RX NCBI\_TaxID=164609;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=22297718; PubMed=12219091;  
 RA Akman L., Yamashita A., Watanabe H., Oshima K., Shiba T., Hattori M.,  
 RA Aksoy S.;  
 RT "Genome sequence of the endocellular obligate symbiont of tsetse  
 RT flies, Wiglesworthia glossinidia."  
 RT Nat. Genet. 32:402-407(2002).  
 DR EMBL: AB063522; BAC24719.1;  
 KW Complete proteome.  
 SQ SEQUENCE 366 AA; 40407 MW; 159E0169E8FC8DED CRC64;

Query Match 65.74% Score 1226.5; DB 16; Length 366;  
 Best Local Similarity 65.98% Pred. No. 1.2e-82;  
 Matches 234; Conservative 62; Mismatches 58; Indels 1; Gaps 1;

QY 7 IQRRKSTRIVGNVPIGDGAPIAVOSMTNRTTVDVATVNOIKALERVADIVR 66  
 DB 7 IQRRKSTRIVGNVPIGDGAPIAVOSMTNRTTVDVATVNOIKALERVADIVR 66  
 QY 67 DAEEFKLIQSVNPLVADIHFDYRIALKAEEGVDCILRNIGNNEIRIM 126  
 DB 67 DAEEFKLIQSVNPLVADIHFDYRIALKAEEGVDCILRNIGNNEIRIM 126  
 QY 67 EAAEFKLIQSVNPLVADIHFDYRIALKAEEGVDCILRNIGNNEIRIM 126  
 DB 67 EAAEFKLIQSVNPLVADIHFDYRIALKAEEGVDCILRNIGNNEIRIM 126  
 QY 127 DKNIPRIGVNGSLKEDLOEKYGEPTPQALLSARHVDHLRNFDOFKVSY 186  
 DB 127 DKNIPRIGVNGSLKEDLOEKYGEPTPQALLSARHVDHLRNFDOFKVSY 186  
 QY 187 LAVESYRLAKQIDOPHLGITTEAGGARSAGVKSALIGLLSGIGDTLAVSL 246  
 DB 187 LAVESYRLAKQIDOPHLGITTEAGGARSAGVKSALIGLLSGIGDTLAVSL 246  
 QY 247 EKVGFSTIRSGINFLIACPTCSROEDVIGTVNALQORLEDTITPMDV 306  
 DB 247 EKVGFSTIRSGINFLIACPTCSROEDVIGTVNALQORLEDTITPMDV 306  
 QY 307 VNGPGEALVSTLGTGNGKSGLYEDGVR-KDRLDNNDMIDOLEARIRAKASQ 360  
 DB 307 VNGPGEALVSTLGTGNGKSGLYEDGVR-KDRLDNNDMIDOLEARIRAKASQ 360  
 QY 361 VNGPGEALVSTLGTGNGKSGLYEDGVR-KDRLDNNDMIDOLEARIRAKASQ 360  
 DB 361 VNGPGEALVSTLGTGNGKSGLYEDGVR-KDRLDNNDMIDOLEARIRAKASQ 360

## RESULT 4

Q8G7Y6 PRELIMINARY: PRT: 403 AA.  
 AC Q8G7Y6;  
 DT 01-MAR-2003 (TREMBlrel, 23, Created)  
 DT 01-MAR-2003 (TREMBlrel, 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel, 23, Last annotation update)  
 DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase 1.

GN ISPG OR BL0098.  
OS Bifidobacterium longum.  
OC Bacteria: Actinobacteria: Actinobacteridae: Bifidobacteriales:  
OC Bifidobacteriaceae: Bifidobacterium.  
OX NCBI\_TaxID=216816;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=NC 2705;  
RX MEDLINE=22294977; PubMed=12381787;  
RA Schell M.A., Karimlantzou M., Snel B., Vilanova D., Berger B.,  
RA Pessi G., Zwalen M.-C., Desiere F., Bork P., Delley M.,  
RA Pridmore R.D., Arigoni F.  
RT "The genome sequence of Bifidobacterium longum reflects its adaptation  
RT to the human gastrointestinal tract."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).  
DR EMBL: AEO14624; FAN233963.1; -.  
SQ Complete proteome.  
KW Sequence 403 AA; 42564 MW; F6F8C3A6153A056 CRC64;

Query Match 43.6%; Score 814.5; DB 16; Length 403;  
Best Local Similarity 46.3%; Pred. No. 4.2e-52;  
Matches 167; Conservative 70; Mismatches 111; Indels 13; Gaps 4;

QY 3 NQAPIQ-RRKSTRIVYGNVPIDGAPLAVQSMNTTPTDVEATVNOIKALERVADIVRY 61  
DB 22 SESPLHRRKSRIRIMVGVGGAGISVQSMNTTPTDVEATVNOIKALERVADIVRY 81  
QY 62 SVFTMAAEAFKLIKQOVNPLVADIHFDYRIALKAEXGVDCRLINPNIENRIRIM-120  
DB 82 AVSODDADLPETCRSPVITADHFOGSKYFOALDAGCAAVRNPNIRKFEVGP 141  
QY 121 VDCARDKNIPRIGVAGSLERDQEKYGEPTPOALLSARHVDHLRLNDOEKVSY 180  
DB 142 ICRATDAGISLRIGVAGSLERDQEKYGEPTPOALLSARHVDHLRLNDOEKVSY 201  
QY 181 KADSVPLAVSYRLKQIDQPHLGTTEAGARSAVKAIGLLSGIDTLRVSL 240  
DB 202 KHHDDVTIMVETRYLRSKGMPLHGTTEAGARSAVKAIGLLSGIDTLRVSL 261  
QY 241 AADPVEIRKGFILKSLRIRGINFACPTCSROEFDVIGTVNALEORLITPMDV 300  
DB 262 SAPPAEAVKYGCKLLEWGLRPRKFDLISPCSGRAVDYQALASNTBELKVTAPIRV 321  
QY 301 SIIGCVNVPGEALVSTLGTGNGKSGLYEDG----VKKRLDN-----NDMIDOLE 349  
DB 322 AVMGCIYNGGEAREADLGVASNGKGOIRIKGVKITVPEDOIIVDLITLTIADIAOME 381  
QY 350 A 350  
DB 382 A 382

## RESULT 5

Q8FP82 PRELIMINARY; PRT; 393 AA.  
AC Q8FP82;  
DT 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Putative aminoglycoside acetyltransferase negative regulator.  
GN CE1903.  
OS Corynebacterium efficiens.  
OC Bacteria: Actinobacteria: Actinobacteridae: Actinomycetales:  
OC Corynebacteriaceae: Corynebacteriaceae: Corynebacterium.  
OX NCBI\_TaxID=152794;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;  
RA Kawarayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,  
RA Ikeo K., Suzuki M., Maehima J., Itoh T., Yamagishi A., Nishio Y.,  
RA Ueda Y., Sugimoto S.;  
RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";  
RT Submitted (May-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL: AP005220; BAC18713.1; -.  
KW Transference: Complete proteome.  
SQ SEQUENCE 393 AA; 41533 MW; A8CC2A98F150AD66 CRC64;

Query Match 40.4%; Score 753; DB 16; Length 393;  
Best Local Similarity 43.7%; Pred. No. 1.4e-47;  
Matches 157; Conservative 67; Mismatches 123; Indels 12; Gaps 3;

QY 9 RKRSTRIVYGNVPIDGAPLAVQSMNTTPTDVEATVNOIKALERVADIVRYPTMDA 68  
DB 20 RKRTRQALMGVSGVSDHPISVQSMNTTPTDVEATVNOIKALERVADIVRYACPKPD 79  
QY 69 AEAFLIKQOVNPLVADIHFDYRIALKAEXGVDCRLINPNIENRIRIMVDCARD 127  
DB 80 AEAFLIKQOVNPLVADIHFDYRIALKAEXGVDCRLINPNIENRIRIMVDCARD 139  
QY 128 KNIPRIGVAGSLERDQEKY-GEPTPOALLSARHVDHLRLNDOEKVSYKASDYF 186  
DB 140 AGIPRIGVAGSLERDQEKY-GEPTPOALLSARHVDHLRLNDOEKVSYKASDYF 199  
QY 187 LAVESYRLAKQIDQPHLGTTEAGARSAVKAIGLLSGIDTLRVSLAADPYE 246  
DB 200 LWEAVARQALAEKCDYPLHGTTEAGARSAVKAIGLLSGIDTLRVSLADPYE 259  
QY 247 EIRKGFILKSLRIRGINFACPTCSROEFDVIGTVNALEORLITPMDVSTIGCV 306  
DB 260 EIVVGDOIILQSLTRKRLKLEIVSCPSGCAQVDYVLAEEVTEGIDGLEVPRLVAVMGCV 319  
QY 307 VNGGEGALVSTLGTGNGKSGLYEDGVKKRLDNDNDMDOLARIRAKASOLDEARRI 365  
DB 320 VNGGEGARDDAGVAGSNGKGOIRIKGVKITVPEDOIIVDLITLTIADIAOME 381

## RESULT 6

Q8EU16 PRELIMINARY; PRT; 362 AA.  
AC Q8EU16;  
DT 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Peptidoglycan acetylation.  
GN MYP9400.  
OS Mycoplasma penetrans.  
OC Bacteria: Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=48227;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=HF-2;  
RX MEDLINE=22354719; PubMed=12466555;  
RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,  
RA Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.,  
RT "The complete genomic sequence of Mycoplasma penetrans, an  
RT intracellular bacterial pathogen in humans."  
RL Nucleic Acids Res. 30:5293-5300(2002).  
DR EMBL: AF004174; BAC44727.1; -.  
KW Complete proteome.  
SQ SEQUENCE 362 AA; 39237 MW; 3F7EB12EF31AE73 CRC64;

Query Match 39.6%; Score 739.5; DB 16; Length 362;  
Best Local Similarity 41.0%; Pred. No. 1.3e-46;  
Matches 146; Conservative 82; Mismatches 127; Indels 1; Gaps 1;

QY 9 RKRSTRIVYGNVPIDGAPLAVQSMNTTPTDVEATVNOIKALERVADIVRYPTMDA 68  
DB 6 RKRTRQALMGVSGVSDHPISVQSMNTTPTDVEATVNOIKALERVADIVRYACPKPD 65  
QY 69 AEAFLIKQOVNPLVADIHFDYRIALKAEXGVDCRLINPNIENRIRIMVDCARD 128  
DB 66 AEAFLIKQOVNPLVADIHFDYRIALKAEXGVDCRLINPNIENRIRIMVDCARD 125  
QY 129 NIPRIGVAGSLERDQEKYGEPTPOALLSARHVDHLRLNDOEKVSYKASDYF 188  
DB 126 NIPRIGVAGSLERDQEKYGEPTPOALLSARHVDHLRLNDOEKVSYKASDYF 184



DR EMBL; AE014468; AAN30676.1; -  
DR TIGR; BR1778; -  
KW Complete proteome.  
SQ SEQUENCE 420 AA; 45029 MW;

SEQUENCE 420 AA; 45029 MW; F1495DB2D73E164C CRC64;

Query Match	27.28;	Score 508;	DB 16;	Length 420;
Best Local Similarity	33.28;	Pred. No. 2.2e-29;		
Matches 128; Conservative	79.	Mismatches 100.	T-adj-1	70

**Dy** 6 P I O R K S T R I Y V G N B D I G D A P I A V O S M T N T R T T D E A V N Q I K A L E R V G A D I Y A V S V P T 65  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
**Db** 12 P P F R R S O V S G V S G V I G V G S A P V V V G S M T N T D A D V S T V A Q V A L H R A G S I Y A I T D R 71

DQ 66 MDAAEAFKLIKQV-----NPLVDADHF-----DY-RALKVAEYGVDCLRINPGENI 112

D0 72 DESAAVPKRIRERLERLGHDPVLVDGFHYGHKLADHPACAEALAY-----RINPENY 126

```
QY      113 G-----NEERIRMYVDCARDKNIPRIIGVNAGSLEKDL-----QEYYGEPYP-----Q 155
        |          :: :| :|:||||| ||::|| :| |
Db      127 GFYDKKKDOPADIVEMAIRYRKPVATGVNNGSSIDPOTIEMTVDNRNOIDPOTICGCTGGTGT 162
```

156 ALLESAMRHVDHDLRNF--DQFKSVKADYFLAVESYRIIAKOIDPLHGTTEAGGA 213

214 RCGAVKSAIGLLSEGIDTLKSLADP---VEIKVGFDDIKSLRIRSGINFIA 269

24 / RRGIVASSAAMGILLQGGIGDTINISLTPEBGDXTRVOVAQELLQTMGRQFVPIAA 306

270 CPTCSROBEDVIGTVNALEQRL-----EDT-----ITPMDSIIGCVVNGPG 311

Db 307 CPGCGR-----TTSTVFQELAQTIQEDIRRNPLMREKYPGEALSVAMGCIYNGP 359

Oy 312 EALVSTICV----TGNKKSGLYEDG 333

```

Db      360 ESKHADIGISLPGTGETPSAPVFDG 385
          | : : : | | : : | |

```

RESULT 10  
O8GZR6  
ID O8GZR6  
PRELIMINARY;  
PRT; 740 AA

DT	01-MAR-2003	(TREMBLrel. 23, Created)
DT	01-MAR-2003	(TREMBLrel. 23, last sequence update)
DT	01-MAR-2003	(TREMBLrel. 23, last sequence update)

DE Geoph.  
05 Lycopersicon esculentum (Tomato).  
06 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta

OC Asteridae; lamids; Solanales; Solanaceae; Solanum.  
NCBI\_TaxID=4081;  
[1]

Queroi J., Imperial S., Boronat A., Rodríguez-Concepción M.;  
"Identification and functional analysis of plant orthologs of the  
Arabidopsis *AtHKT1* gene family"

Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
EMBL: AF435086; AA015447.1; -

SEQUENCE	740 AA;	82230 MM;	8668C9F08078FA05	CRC64;
Query Match	25.28;	Score 471;	DB 10;	Length 740;
Best Local Similarity				

Matches 109; Conservative 45; Mismatches 84; Indels 34; Gaps 5;  
9 RRRSTRIVGVNPIDGAPIAVOSMNTRTDYEAATVNOIKALFRVCAADIVSVSRNDAA 60

69 AAG-----KTRQWNNPDI VASTTNDKQVYI  
84 RRGCTVMGNGVALGSEHNPRIQTMTTDFRDXVAATVEGVMMIADACADIVRTVGSRKE 143

```

144 ADACFEIKNTLVQKNYNIPVLADIHFAPSVALRVAEC-FDXIRVNGNFADRRQFEQLE 202

```

QY 115 -----EEIRRVVDCARDKNIPRIGVNAAGLEKDLQEKYGEPTPALLES 161  
 Db 203 YTEDDYQKLEHEIEETFEPTLVKCRCKYGRAMIGTHNSLDRIIMSYGD-SPRGWESA 261  
 QY 162 MRHVHDLDRINFEQFVSVKASDVLAVESVYLLAKOI-----DQPLHIGTTEGGRSG 216  
 Db 262 FEHARICRKLNDHNFVFSKMASNPVYVQAYKLLVAENYVQGWMDYPLHLGTGEGEGEDG 321  
 QY 217 AVKSAIGLGLLSEGIQDTRLRVSLAADVEEII 248  
 Db 322 RMKSAIGITGLLQDGDIGDPIRVSLEAREEEEL 353

RESULT 11  
08RXG8  
TD 000000  
\*\*\*\*\*

	LAST	LAST DATE
AC	QBRXG8;	
DT	01-JUN-2002 (TREMBLrel. 21, Created)	
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)	

01 01-MAR-2003 (TIEBLrel. 23, last annotation update)  
DE GCPE protein.  
GN AT5G60600.  
OS Arabidopsis thaliana (mustard family)

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons, core eudicots; Rosidae  
OC eurosids II; Brassicales; Brassicaceae; Arabidionsta

```

0A NOBL_ID=3702;
RN [1]
RP SEQUENCE FROM N.A.
NA Nguyen M Karim-Noumane C Contributed by ...

```

RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,

RA Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY081261; AAI91150.1. -

DR InterPro; IPR006705; GcPE.  
DR InterPro; IPR004588; IspC.  
DR Pfam; PF04551; GcPE; 1.  
DR InterPro; IPR006705; GcPE.

SO SEQUENCE 741 AA; 82157 MW; 957B11538FA81585 CRC64;  
Query Match 24.8% Score 463. DP 10. Length 741.

Best Local Similarity 38.6%; Pred. No. 1e-25;  
Matches 105; Conservative 47; Mismatches 86; Indels 34; Gaps 5;

**D**b  
||| : ||| : | | : ||| : | : ||| :  
84 RRKRTVMGNVALSSEHPICIQTMITSDDKITGTVDEVNRIADKGADIVRITVGKKKE 143

09 AEA-----KLIKQVNVNPLVADIHEDYRIALKVAEAGVDDLRINPGNIGNEERIRMYVD 123  
 144 ADACFEIKDKLVOLVNYNIPLVADIHFAPTVALRVAEC-FDKRIVNPGSGFADRRRAQFETID 202

```

124 CARDK-----NIPRLRGVNAAGSLKEDLOERYEPTPOLLESIA 161
      :      :||| :||| :||| :||| :||| :||| :||| :|||
203 YTEDEYQKLEOHIEOVETPLVECKKKYGRARIGTNGSTNRIYMSVYV--SDPQYV 263

```

[illegible]

217 AVKSAIGLILLSEIGDIFLRVSLAADPVEEI 248  
:|||||: || :|||:|||| |||

RESULT 12

09FF59  
09FF59 \*  
09FF59 PRELIMINARY;  
AC 09FF59; PRT; 716 AA.

01-MAR-2001 (Tremblrel. 16) Last sequence update)  
 01-MAR-2003 (Tremblrel. 23) Last annotation update)  
 Gcpe protein.  
 Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_TaxID=3702;  
 [1]  
 SEQUENCE FROM N.A.  
 RC STRAIN=Columbia;  
 RX MEDLINE=97471965; PubMed=9360910;  
 RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,  
 Miyajima N., Tabata S.,  
 "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence  
 features of the 1.6 Mb regions covered by twenty physically assigned  
 pl clones."  
 RL DNA Res. 4:215-230(1997).  
 DR EMBL: AB005246; BAB09833.19  
 DR InterPro: IPR004588; ISPG:  
 DR TIGR00612; ISPG\_Gcpe; 1.  
 SQ SEQUENCE 716 AA; 79817 MW; 28D36C644EB7CAA CRC64;

Query Match 24.83% Score 462; DB 10; Length 716;  
 Best Local Similarity 38.68% Pred. No. 1.2e-25;  
 Matches 105; Conservative 47; Mismatches 86; Indels 34; Gaps 5;

9 RRSSTRIYVGNPIDGAPVAVOSMTNRTTDEATVNOIKALEVAGDIYRVSPYMDA 68  
 59 RRRRTVMVGNVALGSEHPRIOTMTSDTKITGVDEVMRIADKADIVATVQGRKE 118  
 69 AAEF-----KLKQGVNPLVADHFEDYRIALKAVEYGVDCLRINPGNIGNEERIMVVD 123  
 119 AACEEIKKLVQNTNPLVADHFAPVALRVAC-FDKIRVNGFNDRRAQFETID 177  
 124 CARDK-----NIPRIGVNAGSLDLDKKEGEPFOLLTSA 161  
 178 YTEDEYQKELQHEQVFTPLVEKCKRYGRAMRIGTHGSLSDRIMSYGD-SPKMWESA 236  
 162 MRHVHDLRLNFDQFVSKASDVFLAVESYRLAKOI-----DPLHLGITEAGARGSG 216  
 237 FEFARICKKLVQNTNPLVADHFAPVALRVAC-FDKIRVNGFNDRRAQFETID 296  
 217 AVKSAIGLGLLESGIDTLRVSLADPVEEI 248  
 297 RMXSAIGLGLTLDGIDTLRVSLTEPPEEEI 328

RESULT 13  
 08GZRT PRELIMINARY; PRT; 740 AA.  
 ID 08GZRT;  
 AC 08GZRT;  
 DT 01-MAR-2003 (Tremblrel. 23) Created)  
 DT 01-MAR-2003 (Tremblrel. 23) Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23) Last annotation update)  
 Gcpe.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_TaxID=3702;  
 [1]  
 SEQUENCE FROM N.A.  
 RP Querol J., Campos N., Imperial S., Boronat A.,  
 RA Rodriguez-Concepcion M.,  
 "Identification and functional analysis of plant orthologs of the  
 RT Escherichia coli gcpE gene product with a role in plastid isoprenoid  
 biosynthesis."  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF344673; AA015446.19  
 SQ SEQUENCE 740 AA; 82157 MW; 8EF625E9A9C88074 CRC64;

Query Match 24.83% Score 462; DB 10; Length 740;

Best Local Similarity 38.68% Pred. No. 1.2e-25;  
 Matches 105; Conservative 47; Mismatches 86; Indels 34; Gaps 5;  
 9 RRSSTRIYVGNPIDGAPVAVOSMTNRTTDEATVNOIKALEVAGDIYRVSPYMDA 68  
 84 RRRRTVMVGNVALGSEHPRIOTMTSDTKITGVDEVMRIADKADIVATVQGRKE 143  
 69 AAEF-----KLKQGVNPLVADHFEDYRIALKAVEYGVDCLRINPGNIGNEERIMVVD 123  
 144 ADACEEIKKLVQNTNPLVADHFAPVALRVAC-FDKIRVNGFNDRRAQFETID 202  
 124 CARDK-----NIPRIGVNAGSLDLDKKEGEPFOLLTSA 161  
 203 YTEDEYQKELQHEQVFTPLVEKCKRYGRAMRIGTHGSLSDRIMSYGD-SPKMWESA 261  
 162 MRHVHDLRLNFDQFVSKASDVFLAVESYRLAKOI-----DPLHLGITEAGARGSG 216  
 262 FEFARICKKLVQNTNPLVADHFAPVALRVAC-FDKIRVNGFNDRRAQFETID 321  
 217 AVKSAIGLGLLESGIDTLRVSLADPVEEI 248  
 322 RMXSAIGLGLTLDGIDTLRVSLTEPPEEEI 353

RESULT 14  
 08LPQ4 PRELIMINARY; PRT; 741 AA.  
 ID 08LPQ4;  
 AC 08LPQ4;  
 DT 01-OCT-2002 (Tremblrel. 22, Created)  
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 AT960600/mud24.10.  
 DE Arabidopsis thaliana (Mouse-ear cress).  
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_TaxID=3702;  
 [1]  
 SEQUENCE FROM N.A.  
 RP Shinn P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Banh J.,  
 RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,  
 RA Hayashizaki Y., Ishida J., Jones T., Kaniya A., Karlins-Neumann G.,  
 RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,  
 RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,  
 RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,  
 RA Yamamura Y., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A.,  
 RA Ecker J.R.,  
 "Arabidopsis cDNA clones."  
 RT Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 RL EMBL: AY094472; AAM19840.1;  
 DR InterPro: IPR006705; Gcpe.  
 DR InterPro: IPR004588; ISPG:  
 DR Pfam: PF04551; Gcpe; 2.  
 DR TIGR00612; ISPG\_Gcpe; 1.  
 SQ SEQUENCE 741 AA; 82257 MW; BCALD3147BD63ACB CRC64;

Query Match 24.5% Score 457; DB 10; Length 741;  
 Best Local Similarity 38.2% Pred. No. 2.9e-25;  
 Matches 104; Conservative 48; Mismatches 86; Indels 34; Gaps 5;

9 RRSSTRIYVGNPIDGAPVAVOSMTNRTTDEATVNOIKALEVAGDIYRVSPYMDA 68  
 84 RRRRTVMVGNVALGSEHPRIOTMTSDTKITGVDEVMRIADKADIVATVQGRKE 143  
 69 AAEF-----KLKQGVNPLVADHFEDYRIALKAVEYGVDCLRINPGNIGNEERIMVVD 123  
 144 ADACEEIKKLVQNTNPLVADHFAPVALRVAC-FDKIRVNGFNDRRAQFETID 202  
 124 CARDK-----NIPRIGVNAGSLDLDKKEGEPFOLLTSA 161  
 203 YTEDEYQKELQHEQVFTPLVEKCKRYGRAMRIGTHGSLSDRIMSYGD-SPKMWESA 261  
 162 MRHVHDLRLNFDQFVSKASDVFLAVESYRLAKOI-----DPLHLGITEAGARGSG 216



